

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004 13:30:40 ; Search time 2546 Seconds  
(without alignments)  
6214.538 Million cell updates/sec

Title: US-09-914-815A-10

Perfect score: 1 ggaagagagcgctggaac.....cccgagccctactctgcc 651

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST.\*  
1: em\_estda:\*  
2: em\_estdm:\*  
3: em\_estln:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hic:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: em\_gsa\_hum:\*  
18: em\_gsa\_hiv:\*  
19: em\_gsa\_pln:\*  
20: em\_gsa\_vit:\*  
21: em\_gsa\_fun:\*  
22: em\_gsa\_man:\*  
23: em\_gsa\_mus:\*  
24: em\_gsa\_pio:\*  
25: em\_gsa\_rtd:\*  
26: em\_gsa\_pbg:\*  
27: em\_gsa\_vtl:\*  
28: gb\_gsa1:\*  
29: gb\_gsa2:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	407	62.5	467	9	AA463818
2	345.4	53.1	568	9	AA463818 zx67d04.r
3	339	52.1	557	9	AV589535
4	334.6	51.4	541	12	AV656559
					AV656559 AV656559
					BI849231 471975 MA

5	329.6	50.6	434	12	BM289220
6	295.4	45.4	720	11	BQ207465
7	286	43.9	729	13	AK003363
8	284.4	43.7	743	14	CB590125
9	283.2	43.5	485	9	AA033499
10	283	43.5	683	14	BY703620
11	280	43.0	725	13	BQ192887
12	279.2	42.9	758	14	CB321093
13	269.2	41.4	946	14	CB590001
14	261.6	40.2	443	9	AA000766
15	261.6	40.2	462	14	W75342
16	260.2	40.0	343	9	AW315021
17	260.2	40.0	449	9	AI322564
18	250.2	38.4	657	14	CB056845
19	246.6	37.9	339	14	W98365
20	245.8	37.8	666	13	BO206122
21	224.6	34.5	414	14	W08017
22	213.2	32.7	371	13	BY337976
23	195	30.0	348	13	BY338389
24	193.4	29.7	281	13	BO574249
25	191.8	29.5	505	9	AA274230
26	191.2	29.4	341	13	BY338388
27	183.8	28.2	304	13	BY336481
28	183.8	28.2	330	13	BY343164
29	183.8	28.2	330	13	BY343229
30	183.8	28.2	349	13	BY337306
31	183.8	28.2	367	13	BY337975
32	180.6	27.7	346	13	BY340550
33	174	26.7	270	9	AA460254
34	171	26.3	363	13	BY337972
35	167.8	25.8	451	9	AV589534
36	167.4	25.7	752	13	BU442736
37	163	25.0	439	10	BF651486
38	162.8	25.0	488	9	AW213665
39	161.4	24.8	321	14	W09955
40	161.4	24.8	498	10	BE668521
41	155	23.8	445	9	AV665658
42	153.6	23.6	299	13	BY336461
43	124.2	19.1	629	28	AZ361524
44	121.8	18.7	234	9	AA004139
45	121.2	18.6	231	9	AA016631

#### ALIGNMENTS

RESULT 1  
AA463818  
LOCUS  
DEFINITION  
zx67d04.r1 Soares total fetus Mb2H8 9w Homo sapiens cDNA clone  
IMAGE:796519 5' similar to TR:G576631 G576631 14 KDA TRANSMEMBRANE  
PROTEIN. ; mRNA sequence.

ACCESSION  
AA463818  
VERSION  
AA463818.1 GI:2188702

KEYWORDS  
SOURCE  
ORGANISM  
Homo sapiens (human)

REFERENCE  
AUTHORS  
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisler, G., Jost, S.,  
Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B.,  
Schellenberg, K., Stepien, M., Tan, F., Theisling, B., White, Y., Wylie,  
T., Waterston, R. and Wilson, R.

Washu-Merck Est Project 1997

Unpublished

Contact: Willson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810

Email: est@wustl.edu  
This clone is available royalty-free through LNL ; contact the



```

Db      520 ACCCCAGCCCC 531

RESULT 3
LOCUS   AV656559
DEFINITION AV656559 Bos taurus brain fetus Bos taurus cDNA clone E1BR04.9H02
5', mRNA sequence.
ACCESSION AV656559
VERSION   AV656559.1 GI:9924689
KEYWORDS EST.
SOURCE    Bos taurus (cow)
ORGANISM  Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 557)
Takauega,A., Hirotsune,S., Itoh,R., Uchizono,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL  21570554
MEDLINE  11713328
COMMENT   Contact: Yoshikazu Sugimoto
          Animal Genetics Division
          Shikawa Institute of Animal Genetics
          Odokura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
          Tel: 81-248-25-5641
          Fax: 81-248-25-5725
          Email: kazusugi@ccococ.ocn.ne.jp
          Single pass sequencing.
          This clone was obtained from a polyA-deleted cDNA library.

FEATURES
SOURCE
1. 557
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /clone="E1BR04.9H02"
   /tissue_type="brain"
   /dev_stage="fetus"
   /lab_host="DH10B"
   /clone_id="Bos taurus brain fetus"
   /note="Vector: pZ1; Site 1: SalI; Site 2: NotI; Poly A
was deleted from a NotI site"
BASE COUNT      89 a      214 c      158 g      91 t      5 others
ORIGIN
Query Match      52.1%; Score 339; DB 9; Length 557;
Best Local Similarity 83.9%; Pred. No. 3.8e-60;
Matches 407; Conservative 0; Mismatches 69; Indels 9; Gaps 2;

QY      9 CGGCGCTGGAACCATGACACGGGCTATCCCGGAGAGACACCGGGCCCCCAAGGCCCA 68
      12 CGAGCTGGAACCATGACACGGGCTATCCCGGAGAGACACCGGGCCCCCAAGGCCCA 71
      69 GCAAGGCGCG-----GTGCCACACAGGCGCTCACTGGCGGCGCGGACCGCGCGCTC 122
      72 GCAAGGCGGAGGCGGACGCGCGCGCTGAGCGCTGGGGCTTCGGAGACCCCACTTC 131
      123 GAGACCACTGATGTCGTGGGTTCAGACACCGCTTCACTGAATGTCGTGGCTCGGCT 182
      132 GAGACCACTGATGTCGTGGGTTCAGACACCGCTTCACTGAATGTCGTGGCTCGGCT 191
      183 TCCGCGCGCTGCGCTTCACTGAAGGCGCGGAGATCAAGAGTGGTGGTGAAGCTCGAAG 242
      192 TCCGCGCGCTGCGCTTCACTGAAGGCGCGGAGATCAAGAGTGGTGGTGAAGCTCGAAG 251
      243 CGGCGCGCGCTTTGGCTCCAAAGCAGGTCTCAACATCTGGCGCGGAGTGGAGCGC 302
      252 CAGCGCGCGCTGCGCTGAGTCCAAAGCAGGTCTCAACATCTGGCGCGGAGTGGAGCGC 311

```

```

QY      303 TGTGCGCGCACTGCTGCTCTCTGGGGCTGTGGTGTACTGTGTGCTTGACCTTGCGCGCGC 362
      312 TGTGCGCGCGCGCTGCTGCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 371
      363 TGGCCAGAGACTTGGCGCGCTTCTTACAGACCAAGTTGATGAGCGCGGACCTATGACTAC 422
      372 TGGCCAGAGAGCTTGGCGCGCTTCTTACAGACCAAGTTGATGAGCGCGGACCTATGACTAC 431
      423 AGGCTGGGTCCTGATCTGCGGCACTAGCGCCAGACACTGACCCGAGGCTGCGCGCTG 482
      432 AGGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 487
      483 GGGCGC 487
      489 CCCCC 493

Db

QY      483 GGGCGC 487
      489 CCCCC 493

RESULT 4
LOCUS   B1849231
DEFINITION B1849231 541 bp mRNA linear EST 04-OCT-2001
ACCESSION B1849231
VERSION   B1849231.1 GI:15961750
KEYWORDS EST.
SOURCE    Bos taurus (cow)
ORGANISM  Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 541)
Smith,T.P.L., Grose,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casaas,E., Wray,J.E., White,J., Cho,J., Fahrnenkrug,S.C., Bennett,
G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Pertes,G., Holt,I., Karaycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
JOURNAL  21180013
MEDLINE  11282978
COMMENT   Contact: Smith TPL
          USDA, ARS, US Meat Animal Research Center
          PO Box 166, Clay Center, NE 68933-0166, USA
          Tel: 402 762 4366
          Fax: 402 762 4390
          Email: smith@meat.ars.usda.gov
          Single pass sequencing. Bases called and alt trimmed with phred
          v0.980904.e. Vector identified by cross_match with the -minscore 18
          and -mismatch 12 options.
          PCR primers
          FORWARD: AGGAAACAGCTATGACCAT
          BACKWARD: GTTTCCAGTCACGACG
          Plate: 100 row: H column: 8
          Seq primer: ATTAGGTGACACTATAG.

FEATURES
SOURCE
1. 541
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /tissue_type="pooled"
   /lab_host="DH10B"
   /clone_id="MARC 2B0V"
   /note="Vector: pCMV SPOR6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from testis, thymus,
semilendonsus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT      87 a      214 c      151 g      89 t
ORIGIN
Query Match      51.4%; Score 334.6; DB 12; Length 541;
Best Local Similarity 84.6%; Pred. No. 3e-59;
Matches 402; Conservative 0; Mismatches 64; Indels 9; Gaps 2;

```

FEATURES	Source
19	ACCATGAGACACGGGCTATCCCGGAGAGACACCCGGGCCCCACATGCCCAAGACAGCCG- 77
Dd	1 ACCATGAGACAGTGTGTAACCCCGGAGAGACCCCGGGCCCCGAGCCCGCAAGCCGGA 60
Oy	78 -----GTGCCACACAGCCCTCACTAGTGGCGGCCCGGACCCCGCGCTCGAGACACTT 132
Dd	61 CGGCAACGCCCAACAGGCCCTTGACGCTGGGGGCTCCGGAGACCCCACTTCAGACCACTT 120
Oy	133 GATCTGTGGGTGTTCAGACACCCTCTACCTGAATCTGTGTGCTCGGCTTCTGGCGCT 192
Dd	121 GATCTGTGCCGTGTTCACACACCTCTACTGAACCTGTGCTGTCTCGGCTTCTTGGCGCT 180
Oy	193 GGGCTACTCCCAATCAAGGGCCGAGATCAAGAGGTGTGTGTGAACCTGTGAAGGGGCCCGCCG 252
Dd	181 GGGCTACTCCCAATCAAGGGCCGAGACCAAGAGGTGTGTGTGAATCTGTGAAGGAGAGCCCGGCG 240
Oy	253 TTTTGGCTCCAAAGCCCAAGTGTCAACATCTGTGCGCGCGATGTGTGAAGCTGTGTGCCCGC 312
Dd	241 TCTGGGCTCCAAAGCCCAAGTGTCAACATCTGTGCGCGCGATGTGTGTGCGCGCTGTGTGCCCGC 300
Oy	313 ACTGCTGTCTCTGGGGCTGT 372
Dd	301 GGTCTGTCTCTCTGT 360
Oy	373 CTGTCGGCGCTCTCTTCAAGACCAAGTTGT 432
Dd	361 CTGTCGGCGCTCTCTTCAAGACCAAGTTGT 420
Oy	433 CTGATCTGTGGGCACTAGGCCCAAGACACTGACATCCCAAGGCTGTGCTGTGGGGGCC 487
Dd	421 CTGTGCTCTG---TCTGTACCAAGGAGCCCAACCCCAAGGCTCTTACCCAGGCCCGC 472
RESULT 5	
LOCUS	BM289220 434 bp mRNA linear EST 28-DEC-2001
DEFINITION	5307724 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION	BM289220
VERSION	BM289220.1 GI:1798246
KEYWORDS	EST.
SOURCE	Bos taurus (cow)
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidea; Bovidae; Bovinae; Bos.
REFERENCE	1 (bases 1 to 434)
AUTHORS	Smith,T.P.L., Gray,G.W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Caeas,E., Wray,J.E., White,J., Cho,J., Fahrnkung,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitto-McKown,C.G., Pettes,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
TITLE	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL	Genome Res. 11 (4), 626-630 (2001)
MEDLINE	21180013
PUBMED	11282978
COMMENT	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 18 and -mismatch 12 options. PCR Primers FORWARD: AGGAAACAGTATGACCAT BACKWARD: GTTTCAGCTACAGAG Plate: 144 row: M column: 3 Seq primer: ATTAGGACGACTATAG. Location/Qualifiers 1..434

Query Match	Best Local Similarity	88.2%	Pred. No. 3,1e-58;	DB 12;	Length 434;
Matches 372;	Conservative 0;	Mismatches 44;	Indels 6;	Gaps 1;	
20	CCCATGGAACAGCGGCTATCCCGGAGAGACACCCGGGCCCCACGCCCAGCAAGCGG--	77			
1	CCCATGGAACAGCGGCTATCCCGGAGAGACACCCGGGCCCCACGCCCAGCAAGCGG	60			
78	----GTGCGCCACACAGCGCTCAACATGCGCGGCCCCGACCCCGGCTCGAGACCACTTG	133			
61	GCGACAGCCACACAGCGCTCAACATGCGCGGCCCCGACCCCGGCTCGAGACCACTTG	120			
134	ATCTGTCGCTGTTTCAACACCCCTTACCTGATCTGTTTCTCGGCTTCTGAGCGCTG	193			
121	ATCTGTCGCTGTTTCAACACCCCTTACCTGATCTGTTTCTCGGCTTCTGAGCGCTG	180			
194	GCCATCTTCATCAAGCGCGGAGACCAAGTGTGTGTACCTTGGAAGCGGCCCCGCGT	253			
181	GCCATCTTCATCAAGCGCGGAGACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	240			
254	TTTGCGCTCAAGCGGAGCTCAACATCTGCGCGCGGATGTGTGTGTGTGTGTGTGTGT	313			
241	CTGGGCTCAAGCGGAGCTCAACATCTGCGCGCGGATGTGTGTGTGTGTGTGTGTGT	300			
314	CTGCTGCTCTCGGCGCTGT	373			
301	GTCTGCTCTCTCGGCGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	360			
374	TTCGCGCGCTTCTTCAACACCAAGTTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT	433			
361	TTCGCGCGCTTCTTCAACACCAAGTTGATGTGTGTGTGTGTGTGTGTGTGTGTGTGT	420			
434	TG 435				
421	TG 422				

RESULT 6  
B0207465/c

LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
ORGANISM  
SOURCE

REFERENCE  
AUTHORS  
TITLE

JOURNAL  
MEDLINE  
PUBMED

COMMENT

720 bp mRNA linear EST 02-MAY-2002

UI-R-DY1-cno-e-14-0-UI.s1 UI-R-DY1 Rattus norvegicus cDNA clone

UI-R-DY1-cno-e-14-0-UI 3', mRNA sequence.

B0207465 GI:20423930

EST.

Rattus norvegicus (Norway rat)

Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 720)

Bonaldo,W.F., Lennon,G. and Soares,M.B.

Normalization and subtraction: two approaches to facilitate gene  
discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

```

/organism="Rattus norvegicus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/tissue_type="pooled"
/lab_host="DH10B"
/clone_id="MARC 380V"
/notes="Vector: pCMV SPORT6; Site 1: NotI; Site 2: SalI;
library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."

```

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu

The sequence contained an oligo-dT track that was present in the oligonucleotide that was used to prime the synthesis of first strand cDNA and therefore this may represent a bonafide poly A tail. The sequence tag present in the cDNA between the NotI site and the oligo-dT track served to verify it as a clone from the normalized cartilage library cDNA library preparation. M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com)  
Seq primer: M13 Forward  
POLYA=yes

# FEATURES

Location/Qualifiers  
1. 720  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DY1-cno-e-14-0-UI"  
/tissue\_type="Cartilage"  
/dev\_stage="ADULT"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-DY1"  
/note="Organ: Femur and Tibia; Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; UI-R-DY1 is a normalized cDNA library containing the following tissue(s): Rat Cartilage from Femur and Tibia. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTAATGACG. The Rat Cartilage tissue was provided by Dr Jeff Stevens at the University of Iowa.  
TAG\_LIB=UI-R-DY1  
TAG\_TISSUE=cartilage  
TAG\_SEQ=CTAATGACG"

BASE COUNT 153 a 173 c 226 g 163 t 5 others  
Query Match 45.4%; Score 295.4; DB 13; Length 720;  
Best Local Similarity 79.8%; Pred. No. 4e-51;  
Matches 360; Conservative 0; Mismatches 85; Indels 6; Gaps 1;

QY 1 GGAAGAGACGGCGCTGGAACCCATGACACGCGCTATCCCGCGAGAGACCCGCGCC 60  
DB 720 GGAAGAGACGGCGCTGGAACCCATGACACCTTATCNCCTGAGAGACCCGCGCC 661  
QY 61 CAGGCCACGACGACGCGCT-----GCCACACAGCCCTCACTAGCGCGCCGACACC 114  
DB 660 ATCATCCCGAAGGNTATGTCAGGCCACACGCGCCCTCTGTGAGACACCCGCGCC 601  
QY 115 CCGGCTCGAGACCACTTGATCTGTGCGGTGTGACGACCCCTACTGTAATCTGTGTTG 174  
DB 600 TAGCCACGAGACCACTTGCTCTGTGCTGTCTTGAAGACGCTGTAAGTCTGTGCTG 541  
QY 175 CTTGCGCTTCTGCGCGCTGACCTCATCAACAGCCCGAGATCAAGAGTGTGTGTA 234  
DB 540 CTTGCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481  
QY 235 CTTGCGAGCGCGCGCGCTTGTGCTCCAAAGCCAGTGTCAACAATCTCTGCGCGGAT 294  
DB 480 CTTGAGAGCTGCAAGAGATGTGCTCCAAAGCCAAATGTCTCAACAATCTCTGCTGCAAT 421  
QY 295 GTGACACCTGGTGCAGCACTGTGCTCTGCGGCGCTGTGAGTGACTGTGCGCCCTGACCT 354

DB 420 GTGACACCTGTGTCGCCCACTTGCTGCTCTGCGGACTGTGTGATCTGCGCGCTTGCACT 361  
QY 355 GCGCCGCTGTCGCCCAAGACTGTGCGCGCTTCTTACAGACCAAGTTGATGACCGGACTA 414  
DB 360 GTCCCGGCTAGCCAAAGACTGTGCGCGCTTCTTACAGACCAAGTTGATGAGAGACTA 301  
QY 415 TACTGACAGGCTGCGGTCTGATCTGAGGCA 445  
DB 300 TAACTGAGAGTTCTGAGCCTGTCTGACCA 270

## RESULT 7

AK003363

## LOCUS

DEFINITION Mus musculus 18-day embryo whole body cDNA, RIKEN full-length enriched library, clone:111003J06 product:haemopoiesis related membrane protein 1, full insert sequence.

## ACCESSION

AK003363.1 GI:12833982  
VERSION AK003363  
KEYWORDS HTC, CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

## REFERENCE

1 Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999)

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000)

3 20499374  
MEDLINE 11042159  
PUBMED

## TITLE

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

AUTHORS

REFERENCE

Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Komno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishii, K., Kiyosawa, H., Komno, S., Yamamoto, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadoya, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glessi, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Staudt, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Balderelli, P., Barsh, G., Blake, J., Botfield, D., Boyle, C., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bul, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, C., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzaletti, J., Mombert, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whitaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, H., Kohsaki, S., and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001)

```

MEDLINE
PUBMED
12108660
11217851
5
REFERENCE
AUTHORS
TITLE
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analyses of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 739)
JOURNAL
REFERENCE
AUTHORS
Adechi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arikawa,T., Bono,H., Caminici,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hiraoka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,Y., Kojima,Y., Komio,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Muramatsu,M. and Hayashizaki,Y.
Direct Submission
Submitted (10-JUN-2000) Yoshinobu Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-reseq@riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5'-GAGAGAGACGCGCCGCAACCTGAGATTTTCTTTTCTTTN 3'], cDNA was
prepared by using trihalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adaptor of sequence[5'-
GAGAGAGACGAGATCGACGACGATCAATTAATTAACCCCCCC 3']. cDNA was
cleaved with XhoI and ScaI. Cloning sites, 5' end: ScaI; 3' end:
XhoI. Host: SOLR.
FEATURES
Source
Location/Qualifiers
1..739
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="FANTOM DB:111000306"
/db_xref="MGI:1906873"
/db_xref="taxon:10090"
/clone="111000306"
/tissue_type="whole body"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18-day embryo"
62..466
/misc_feature
/note="haemopoiesis related membrane protein 1
(MGI:MGI:1934923, GB/AJ009781, evidence: BLASTN, 99%,
match=469)
putative"
/db_xref="MGI:1934923"
721..726
/note="putative"
739
/note="putative"
BASE COUNT
160 a 228 c 186 g 165 t
ORIGIN
Query Match 43.9%; Score 286; DB 11; Length 739;
Best Local Similarity 68.1%; Pred. No. 3.5e-49;
Matches 436; Conservative 0; Mismatches 185; Indels 19; Gaps 2
1 GGAAGAGACGCGCTGGAACCATGACACGCGCTGATCCGCGGAGAGACACCGGAGCC 60

```

Db	Accession	Version	Source	Organism	Reference Authors Title Journal	Comment
Db	40		GGAAGACACGGCGGTGGAACCCATGACACTTCAATATCCCGTAGAGACCCCGGGCTCC			
QY	61		CAGGCCAGCAGAGCCGGT-----GCCACACAGCCCTTCACTGGCGGCCGACACC			
Db	100		ATCATCCCGAAGCGTATGCTGACGCCACACAGCCCTCTCCATGGGAAACACTGGAGCC			
QY	115		CCGCGCCCGAGACCACTGTAATCTGTGTGGAGTTTACAGACCCCTTCACTGTAATCTGTGTG			
Db	160		TACACCAAGAGATACATGCTCTGTGTCTGTCTTACAGACATGTACCTGAATCTGTGTG			
QY	175		CTGCGCTTCCTGGCGCTGGCTCTAATCAATCAAGCCCTGAGATCAAGAGGTGTGTGTGA			
Db	220		CTTTGATCTCTGGCGCTGGCTCACTGTCTCAAGCCCGAAGACTGAAGATGGCTGGGAA			
QY	235		CTTGAAGCGCGCCCGGGGCTTTGGCTCCAAAGCCAAAGTGTACAAACATCTGGCGGGAT			
Db	280		CTTGAAGCGCTGAAGGAGTATGGCTCCAAAGCCAAAGTGTACAAACATCTGGCTGCAAT			
QY	295		GTGACACGCTGTGGCGCCGCTGCTGCTCTGGGCGTGGTGTGACTGTGCTTGCACCT			
Db	340		GTGGAATTTGGTGGCCCCCATTTGCTGCTCTGGAGCTGGTGTGACTGGCGCTTGACCT			
QY	355		GGCGCCGCTGGCCAGGACTCTGCGCCCTTTTACGACCAAGTTTGTATGACGGGACTA			
Db	400		GTCCAAATTTAGCCAAAGACTCTGGGCGCTTTCTTAGACCAAGTTTGTATGAGGAGACTA			
QY	415		TGACTGACAGGCTGGGCTGTGATCTGGGAGCACTAGCCCAAGACACTAACCCAGGCTGC			
Db	460		TAACTAAGATTTCCGAGCTG-----TCCCTGAACCGAGGACAAACATGTCA			
QY	475		TGCCCTTGGGGCCCAATATGACTGCCCGGAGCCTGGCCCTCTTCTGTGGGCTTCAT			
Db	507		GGTACGCTGTGGCCAAACACAGCTCTGGGGTGTACAGCTATATAGCACTTCATCC			
QY	535		CCCGGCCCAATCTGATCTGGGGCCCTCCAGCCCAACATGGGACCTTAAGCTGAAC			
Db	567		TGCCAGCCGTGACTGACTGAGAGCGGGCTCTTGTCCAGATGTATTGATGCTGCC			
QY	595		AGTCAGACCCCGGGGCTTCAACCTTAACCCGAGATTGCC			
Db	627		ACTCAGATCCCTCAGACTTAACCTTAACCTTAGAGGGTCC			
RESULT 8						
CB590125						
LOCUS						
DEFINITION						
ACCESSION						
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
COMMENT						

source

1. .743  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30287396"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: pCMV-Sport6.1.ccdB. Site 1: EcoRV, Site 2: NotI; Normalized, full-length enriched library from pool of mouse embryonic limb, maxilla and mandible, embryonic day 17.5, 18.5 and newborn (mandible 5, 4 and 1 limb and jaw equivalents from respective days). Cloned directionally. oligo-dT primed (5'-GACTAGTTTCAAGTCGAGCGGCGCCGCTTT)15-3'. Size selected for the >1kb fragments, average insert size 1.2 kb. Normalization to Cot 7.5. Tissue contributed by David Rowe; library constructed by ResGen, Invitrogen Corp. Note: this is a NIH MGC library."

BASE COUNT 167 a 226 c 176 g 162 t 12 others

ORIGIN

Query Match 43.7%; Score 284.4; DB 14; Length 743;  
Best Local Similarity 68.0%; Pred. No. 7.6e-49;  
Matches 435; Conservative 0; Mismatches 186; Indels 19; Gaps 2;

QY 1 GGAAAGACGGCGCTGGAACCCATGAGACCGCGCTATCCCGAGAGACCCCGGCCCC 60  
Db 16 GGAAAGACCGCGCTGGAACCCATGAGACCTTCATATCCCGTGAAGACCCCGGCGCTCC 75  
QY 61 CAGCGCCAGACGAGCGCGT-----GCCACAGACCGCTCACTGGCGCGCCGACGCC 114  
Db 76 ATCATCCCGAGAGCTGATGCTGACAGCCACAGCCCTCTCCATGGAACACTGGCGCC 135  
QY 115 CCGCGCTGAGACCACTGATCTGATGCTGATGAGACCCCTTACCTGATCTGATG 174  
Db 136 TACACACAGAGATCAATGCTGCTGCTGCTTACAGACATGATACCTGATCTGCTG 195  
QY 175 CCGCGCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234  
Db 196 CTTGGAATCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255  
QY 235 CCGGAGACCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 294  
Db 256 CTTGGAAGCTGAG 315  
QY 295 GTGAGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 354  
Db 316 GTGAGACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 375  
QY 355 GCGCGCGCTGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414  
Db 376 GTCCAAATTAAGCAAAAGCTGCGGCTTCTTCAAGCAAAAGTTATGAGAGAGACTA 435  
QY 415 TGACTGACAGGCTGAGCTGATCTGAGGAGCACTGAGCCCAAGCACTGAGCCAGGCTGC 474  
Db 436 TAACTAAGAGTCCGAGCCTG-----TCCCTGAACGAGAGCAAAACATGTC 482  
QY 475 TGCCCTGAGGAGCCCAATAGTACTCCCGAGAGCTGAGCCCTCTTCTGAGGAGCTCAT 534  
Db 483 GGTAGCTGAGCCCAACAGAGCTCTGAGAGTTACAGCTTATAGCACTGATGCC 542  
QY 535 CCGTGGCCCACTCTGATCTGAGGAGCTTCAAGCCCAACATGAGGAGCTTAAGGCTAACC 594  
Db 543 TGCCCAAGCCGCTGAGCTGAGAGGCGCGCTCTTGTTCAGATGTATATTATGATGCTGCC 602  
QY 595 AGTCAGACCCCGGAGGCTTCAACCTTAACCGAGAGGTTCCC 634  
Db 603 ACTCAGATCCCTCAGACTTAATTATTAACCTTAAGGAGGCTCC 642

RESULT 9  
AA033499

LOCUS AA033499 485 bp mRNA linear EST 22-AUG-1996  
DEFINITION m144b03.r1 Soares mouse embryo NbM213.5 14.5 Mus musculus cDNA  
IMAGE:465349.5' similar to SW:INB RAT P26376  
ACCESSION AA033499  
VERSION AA033499.1 GI:1504910  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE  
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellendberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wyllie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mousees@watson.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.lnl.gov) for further information.  
MGI:280165  
Seq primer: -28M13 rev2 from Amerisham  
High quality sequence stop: 404.  
Location/Qualifiers  
1. .485  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:465349"  
/sex="unknown"  
/tissue\_type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
/clone\_lib="Soares mouse embryo NbM213.5 14.5"  
/note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAG T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTR73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 108 a 142 c 133 g 102 t

ORIGIN

Query Match 43.5%; Score 283.2; DB 9; Length 485;  
Best Local Similarity 80.5%; Pred. No. 1.2e-48;  
Matches 346; Conservative 0; Mismatches 78; Indels 6; Gaps 1;

QY 1 GGAAAGACGGCGCTGGAACCCATGAGACCGCGCTATCCCGAGAGACCCCGGCCCC 60  
Db 51 GGAAAGACGGCGCTGGAACCCATGAGACCTTCATATCCCGTGAAGACCCCGGCGCTCC 110  
QY 61 CAGCGCCAGACGAGCGCGT-----GCCACAGACCGCTCACTGGCGCGCCCGACGCC 114  
Db 111 ATCATCCCGAGAGCTGATGCTGACAGCCCAAGCGCTCTTCCATGAGAAACACTGGGCC 170  
QY 115 CCGCGCTGAGACCACTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 174  
Db 171 TACACACAGAGATCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 230



Qy	175	CCTGGGCTTCTGGGGCTGGCCCTTACATCAAGGCCAGATCAAGAAAGTGGTGA	234
Db	231	CCTTGATCTCTGGGGCTGGTCCATCTGTCAAGGCCAGAGACAGAAAGATGGCTGGAA	290
Qy	235	CCTGGAAACGGCCCGCGCTTTTGGCTCCAAAGCCAAAGTGTCAACATCTGGCCGCAT	294
Db	291	CTTGAGAGCTCCAAAGGCAAGTATGGCTCCAAAGCCAAAGTGTCTCAACATCTGGCTGCAAT	350
Qy	295	GTGACGCTGGTGGCCGCCACTGCTCTCCCTGGGGGCTGGTGGAGACTGGAGTGGCCCTGCACT	354
Db	351	GTGACATTTGGTGGCCCAATGTCTGTCTCTGGGAATGGTGGTGACTGGGGCCTTTGGACCT	410
Qy	355	GGCCCGGCTGGCCAAAGACTGTGCCGCTTCTTTCAGCACAAGATTGATGACGCGACTA	414
Db	411	GTCCAAAGTACCAAAAGACTGCGCGCTTTCTTCAGCACCAAGTTGATGAGGAAGACTA	470
Qy	415	TGACTGACAG	424
Db	471	TAACTAAGAG	480

RESULT 10	BY703620	LOCUS	DEFINITION
BY703620	683 bp	mRNA	linear
BY703620	RIKEN full-length enriched, 18-day embryo/whole body Mus musculus cDNA clone 1110003J06 5', mRNA sequence.		EST 16-DEC-2002

ACCESSION	BY703620
VERSION	BY703620.1
DATE	GI:27114731

SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus

## REFERENCE AUTHORS

TITLE  
 Analysis of the mouse transcriptome based on functional annotation  
 of 60,770 full-length cDNAs  
 NATURE 420, 563-565 (2002)

PUBMED  
COMMENT

Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Shukho-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: [genome-res@sc.riken.go.jp](mailto:genome-res@sc.riken.go.jp),  
URL: <http://genome.gsc.riken.go.jp/>  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda,  
S., Hashizume, M., Hayashida, K., Hirozane, T., Hori, F., Imotani, K.,  
Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno,  
H., Koya, S., Miyazaki, A., Murate, M., Nakamura, N., Nomura, K.,  
Numasaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sato, H.,  
Sasakhi, D., Sato, K., Shibata, K., Shiraki, T., Tgami, M., Takeda, Y.,  
Waki, K., Watanishi, A., Muramatsu, M. and Hayashizaki, Y. Direct  
Submission  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: a real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (<http://genome.gsc.riken.go.jp>) for  
further details.

## FEATURES

### Source

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1110003706"
/risue_type="whole body"
/dev_stage="18-day embryo"
/clone_lib="RIKEN full-length enriched, 18-day embryo
whole body"
BASE COUNT      150 a      205 c      174 g      154 t
ORIGIN

```

Query Match	43.5%;	Score 283;	DB 14;	Length 683;
Best Local Similarity	79.4%;	Pred. No. 1.4e-48;		
Matches 350;	Conservative	0;	Mismatches 85;	Indels 6;
			Gaps	1
QY	1	GGAAAGACGGGGCTGTGAACCCATATGACAACGGCGTATCTCCCGCAGAGACACCCGGGGCCCC	60	
DB	40	GGAAAGACAGGGGGCTGTGAACCCATATGACAATCTATCTCCCGTAGAGACCCCCGGGGCTCC	99	
OY	61	CACGCCACGAAGGCGCGT-----GCCACACAGCCCTCACATGCGGGCCCCGACCC	114	
DB	100	ATATCTCCGCAAGGCTGATGTCTGACAGCCACACAGCCCTCTCCANTGGGAACACTGGCCCC	159	
OY	115	CCCGCCTCGAACACACTTGTATTTGGTGGGTTCACACCTCTTACTGTAATCTGTGTG	174	
DB	160	TACACCAACAGATCAATGCTCTGTCTGTCTTACAGACAGATGACTCGAATCTGTCTG	219	
OY	175	CCTCGGCTTCTGAGCGCTGAGCTTAATCCATCAAGGCCGAGATCAGAAGGTGTGTGTGA	234	
DB	220	CTTTGATTTCTGGCGCTGTGTCACTGTGTGTAAGGCCGAGACCTAGAAATGCTGGAA	279	
OY	235	CTGTGAAGGGGCCCGCGCTTTTGGCTCCAAAGCCAGTGTCTACAACTCTGGGCCGGAT	294	
DB	280	CTTGAAGGCTGCAAGGCAAGATATGCTCTCCAAAGCCAAAGTCTACAACTCTGGGTGAAAT	339	
OY	295	GTGAGCGCTGTGTCGGCCACTGCTGCTCTCGGGCTGTGTGTGACTGTGTCCTTGCACT	354	
DB	340	GTGGAATTTGTGGCCCCCATTTCTCTCTCGGAACTGGTGTGTGACTGGCGCTTGCACT	399	
OY	355	GGCGCCGGCTGGCCAAAGACTCTGCGGCTTTTACAGACCAAGTTTATATACGGGACTA	414	
DB	400	GTCAAGTTAGCCAAAGACTCTGCGGCTTTTCTTACAGACCAAGTTTATATAGAGAGACTA	459	



QY 415 TGACTGACAGGCTGGTCTTG 435  
DB 460 TAACTAGAGTTCGAGCCTG 480

RESULT 11  
B0192887/c  
LOCUS B0192887 725 bp mRNA linear EST 30-APR-2002  
DEFINITION UI-R-DRI-clk-m-07-0-UI-s1 UI-R-DRI Rattus norvegicus cDNA clone  
B0192887  
UI-R-DRI-clk-m-07-0-UI 3', mRNA sequence.  
B0192887.1 GI:20368438  
EST.  
SOURCE Rattus norvegicus (Norway rat)  
ORGANISM Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 725)  
Bonaldo, M.F., Lennon, G. and Soares, M.B.  
Normalization and subtraction: two approaches to facilitate gene  
discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL MEDLINE 97044477  
PUBMED 8889548  
COMMENT  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
The sequence contained an oligo-dT track that was present in the  
oligonucleotide that was used to prime the synthesis of first  
strand cDNA and therefore this may represent a bonafide poly A  
tail. The sequence tag present in the cDNA between the NotI site  
and the oligo-dT track served to verify it as a clone from the  
normalized osteoblast library cDNA library Preparation: M.B. Soares  
Lab Clone distribution: Clones will be available through Research  
Genetics (www.regen.com)  
Seq primer: M13 Forward  
POLYA=Yes.

FEATURES  
source  
Location/Qualifiers  
1..725  
/organism="Rattus norvegicus"  
/mol\_type="mRNA"  
/strain="Sprague-Dawley"  
/db\_xref="taxon:10116"  
/clone="UI-R-DRI-clk-m-07-0-UI"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/clone\_lib="UI-R-DRI"  
/note="Vector: pTZ19-DRI-Pac (Pharmacia) with a modified  
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-DRI  
library is a normalized Rat Osteoblast library (nRO)  
constructed in pTZ19 vector according to the procedure  
described by Bonaldo, Lennon & Soares (Normalization and  
Subtraction: Two Approaches to Facilitate Gene Discovery.  
Genome Research 6: 791-806, 1996). The oligonucleotide  
used to prime first strand synthesis contained the  
sequence tag AAGATATCA between the Not I cloning site and  
dri8 stretch. The Rat Osteoblast tissue was provided by  
Lien & Stein of the University of Massachusetts Medical  
School.  
TAG LIB=UI-R-DRI  
TAG\_TISSUE=osteoblast  
TAG\_SEQ=AAGATATCA"

BASE COUNT 155 a 175 c 228 g 165 t 2 others

ORIGIN  
Query Match 43.0%; Score 280; DB 13; Length 725;  
Best Local Similarity 79.4%; Pred. No. 6.1e-48;

Matches 358; Conservative 0; Mismatches 86; Indels 7; Gaps 2;  
QY 1 GGAAAGACGGCGCTGGAACCCATGGAACGGCGATATCCCGGAGACACCCGGGCCC 60  
DB 719 GGAAAGACGGCGCTGGAACCCATGGAACCTTATATCCCGGAGACACCCGGGCCC 660  
QY 61 CACGCCACGACGCGCGT-----GCCACACAGCCCTTCACCTGGCGGCCCCGACCC 114  
DB 659 ATCATCCCCCAGGCTGATGATGACACCCACGGGCTCTGTGTG-GACACCGGCCC 601  
QY 115 CCGGCTCGAGACCACTTATGATGCGGATGTTACACCCCTTACTGATCTGTTG 174  
DB 600 TACGCCACGACCAATATGCTCTGCTCTTACACAGCTGATCTGATCTGTCTG 541  
QY 175 CCTGCGCTTCTGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234  
DB 540 CTTGTTTCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 481  
QY 235 CTTGGAAGCGCGCGCGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 234  
DB 480 CTTGGAAGCGCGCGCGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 421  
QY 295 GTGAGACGTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 354  
DB 420 GTGAGACGTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
QY 355 GGCCCGCTGCGCGCAAGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 414  
DB 360 GTCCCGCTGCGCGCAAGACCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 361  
QY 415 TGACTGACAGGCTGGTCTGATCTGGGGCA 445  
DB 300 TAACTGAGAGTTCGAGCCTGCTGTAACCA 270

RESULT 12  
CB321093 738 bp mRNA linear EST 04-MAR-2003  
LOCUS CB321093  
DEFINITION AGENCOURT 12238330 NIH MGC 136 Mus musculus cDNA clone  
CB321093  
IMAGE:30288864 5', mRNA sequence.  
ACCESSION CB321093  
VERSION CB321093.1 GI:28845328  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 738)  
NIH-MGC http://mgi.mgi.nhl.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. David Rowe  
cDNA Library Preparation: Invitrogen Corp  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: NDAM327 row: a column: 01  
High quality sequence stop: 592.  
Location/Qualifiers  
1..738  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="IMAGE:30288864"  
/tissue\_type="embryonic limb, maxilla and mandible"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH MGC 136"  
/note="Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:  
NotI; Normalized, full-length enriched library from pool

FEATURES  
source



OY	495	GATCTCCCGGAGCCTGGGCCCTCTTGTGGGGGCCTTCATCCCTGCCCCCATCTGATCT	554
Dd	468	AGCTCTGGGAGTTACACACTTACTATTAGCACTTGATCTCCGCCAGCGCTGAACTCAGA	527
OY	555	GGGGCCCTCCAGCCCCAACATGGGACCCTAAGGCTGAACCACTCAGACCCGGGGCTTCC	614
Dd	528	GGCGGCCCCTCTTGTCACGATGTGTATTGATGCTGCCCACTGAGATCCCTCAGACTTA	587
OY	615	ACCTAACCAGAGATTCCC	634
Dd	588	ACTTATACTTAGAGGCTCC	607

RESULT 14

LOCUS	AA000766	
DEFINITION	mg8la01.r1 Soares mouse embryo NDME13.5 14.5 Mus musculus cDNA clone IMAGE:426024 5' similar to SW:INIB PAT P26376 INTERFERON-INDUCIBLE PROTEIN. [1] ; mRNA sequence.	
ACCESSION	AA000766	
VERSION	AA000766.1 GI:1436846	
KEYWORDS	EST.	
SOURCE	Mus musculus	
ORGANISM	Mus musculus (house mouse)	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scuriognathi; Muridae; Murinae; Mus.	
AUTHORS	Matra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geiscl,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,B., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.	
TITLE	The WashU-HMI Mouse EST Project	
JOURNAL	Unpublished	
COMMENT	Contact: Matra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINTL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MGJ:260576 Seq primer: ETPRimer High quality sequence stop: 354. Location/Qualifiers	
FEATURES	source	
	1..443	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/strain="C57BL/6J"	
	/db_xref="taxon:10090"	
	/clone="IMAGE:426024"	
	/sex="unknown"	
	/tissue.type="embryo"	
	/dev_stage="13.5-14.5dpc total fetus"	
	/lab_host="DH10B"	
	/clone_idb="Soares mouse embryo NDME13.5 14.5"	
	/note="(vector: pRTD-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5], TGTACCAATCTGAAGTGGGCGGCGGAGAAATTTTTTTTTTTTTTTTTTTTTTTT	
	T 3 ], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ] ; double-stranded cDNA was ligated tc	
	Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRTD vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Felima Bonaldo."	

BASE COUNT

96 a	135 C	115 G	97 C
------	-------	-------	------

ORIGIN

	Query Match	Local Similarity	40.2%;	Score 261.6;	DB 9;	Length 443;
	Best Local Similarity	78.5%;	Pred. No. 3,4e-44;			
	Matches 328;	Conservative 0;	Mismatches 84;	Indels 6;	Gaps 1;	
QY	24	TGGACACGCGGTATCCCGCGAGGACACCCGGGCCCCCAGCCGACGACGAGCCGGT---	79			
Db	3	TGGACATTATATCCCCGATGATACCCCGGGCTTCATATCCCGCAGGCTATGCTG	62			
QY	80	--GCCACACAGCCCTCACCTGGCGGCCCCCGACCCCGGCTCGAGACCACTTATCT	137			
Db	63	CAGCCACACAGCCCTTTCATGGGAACCTGGCCCTTACACACGAGATCAATGCTCT	122			
QY	138	GGTGGATTTACAGACCCCTCACTCTGATCTGTGTCCTTCGCTTCGCGCTGACCT	197			
Db	123	GGTGTCTTCAGACGATGACTGATCTGTGCTGCTGCTGATTCCTGGCGCTGGTCC	182			
QY	198	ACTTCATCAAGGCCCGGAGATCAAGAGTGTGTGACCTGGAAAGCGGCCGGGGTTTG	257			
Db	183	ACTGTTCAGAGGCCCGGACCAAGATGTGCTGGAACTTGGAGCTGCAAGGACGATATG	242			
QY	258	GCTCAAGCAAGAGTGCTACAAACATCTTGCGCGGATGTGACCGTGGTGCGCACATGC	317			
Db	243	GCTCCAAAGCCAAGTCTTACAAATCTTGCTGTAATGTGAACTTGGTGGCCCCATTGC	302			
QY	318	TGCTCTGGGCGTGTGTGATCTGTGTCCTTCGACCTGGCCCGGCTGGCCAAAGACTTGG	377			
Db	303	TGCTCTGGGACGTGTGTGATCTGGCGCCCTTTGCACCTGTCCAAATGTAGCCAAAGACTTGG	362			
QY	378	CCGCGCTCTTCAGACCAAGTTGATGACGGGACCTATGACTGACAGGCTGGGTCTTG	435			
Db	363	CGGCTTCTTCAGACCAAGTTGATGAGAGGACTTATAGAGATGTTCCGAGCTTG	420			

RESULT	15
LOCUS	W75342
DEFINITION	m49n03.r1 Soares mouse embryo NMME3.5 14.5 Mus musculus cDNA clone IMAGE390869.5 similar to SW:INIB_RAT P26376 INTERFERON-INDUCIBLE PROTEIN. [1] ; mRNA sequence.
ACCESSION	W75342
VERSION	W75342.1 GI:1385557
KEYWORDS	EST.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Scurionathi; Muridae; Murinae; Mus. 1 (bases 1 to 462)
AUTHORS	Marrin,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Gaisel,S., Kuchaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theisinger,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
TITLE	The WashU-HIMI Mouse EST Project
JOURNAL	Unpublished
COMMENT	Contact: Marrin M/Mouse EST Project WashU-HIMI Mouse EST Project Washington University School of Medicine# 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through INM; contact the IMAG Consortium ( <a href="mailto:info@image.liml.gov">info@image.liml.gov</a> ) for further information. MWI:242701 Seq primer: mob.RGGA+ET High quality sequence stop: 343. Location/Qualifiers 1..462 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="taxon:10090"
FEATURES	
SOURCE	

```
/clone="IMAGE:390869"  
/sex="unknown"  
/tissue_type="embryo"  
/dev stage="13.5-14.5dpc total fetus"  
/lab host="DH10B"  
/clone.lib="Soares mouse embryo NMME13.5.14.5"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified  
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA  
was primed with a Not I - oligo (dT) primer [5',  
TGTTACCAATCTGAAGTGGAGCGCGCCGGAATTTTCTTTTCTTTT  
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2  
14.5dpc embryos (total RNA provided by Minoru Ko, Wayne  
State Univ., from 2); double-stranded cDNA was ligated to  
Eco RI adaptors (Pharmacia), digested with Not I and  
cloned into the Not I and Eco RI sites of the modified  
pT73 vector. Library went through one round of  
normalization, and was constructed by Bento Soares and  
M.Fatima Bonaldo."
```

```
BASE COUNT      101 a      141 c      123 g      97 t  
ORIGIN
```

```
Query Match      40.2%; Score 261.6; DB 14; Length 462;  
Best Local Similarity 79.2%; Pred. No. 3.4e-44;  
Matches 350; Conservative 0; Mismatches 84; Indels 8; Gaps 3;
```

```
OY      1  GGAAGACGCGGCTGTGAACCCATGACACGCGGTATCCCGCGAGACACCCGCGCC 60  
      5  GGAAGACGCGGCTGTGAACCCATGACACCTTATATCCCGTGA-GACCCCGCGCTCC 63  
OY      61  CACGCGCAGCAAGCGCGT-----GCCACACAGCCCTCAACCTGGCGCGCCGCAACC 114  
      64  ATCATCCGCGAGGCTGATGCTGACAGCCACAGCCCTCTCCATGGAACACCTGGGCC 123  
OY      115  CCCGCTCGAGACCACTGATCTGTGCGATTCAGACCCCTTACTGAACTGTGTG 174  
      124  TACACCAAGATACATGCTGTGCTGTCTTACAGACGATACCTGAATCTGTGCTG 183  
OY      175  CCTGGCTTCTGGGCTGTGCTTCAATCAAGGCCGAGATCGAAGAGTGTGTGA 234  
      184  CTTGGATTCTGGGCTGTGCTCACTGTCAAGGCCGAGACGAGATGGCTGGGAA 243  
OY      235  CTTGAGAGGCGCGCGCTTTTGGCTCCAA-GCGAAGTCTACAAACATCTGGCGCGGA 293  
      244  CTTGAGGCTGACAGGCTGATGCTCCAAAGGCCAAGTCTACAAACATCTGGCTGCA 303  
OY      294  TGTGACGCTGTGCTGCGCACTGTCTCTGGGCTGTGTGACTGTGCTCTGCACC 353  
      304  TGTGACATTTGGTGTCCCATTTGCTGCTCTGGGACTGTGTGACTGGCGCTTGACC 363  
OY      354  TGGCCCGGCTGGCCAGAGACTGTGCGCTTCTTACAGACCAAGTTGATGACGGGACT 413  
      364  TGTCCAGCTTAGCCAAAGACTCTGGGCTTCTTACAGACCAAGTTGATGAGGAGACT 423  
OY      414  ATGACTGACAGGCTGGTCTG 435  
      424  ATAACTAAGATTCCGAGCTG 445  
DB
```

```
Search completed: January 31, 2004, 15:13:27  
Job time : 2556 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 13:38:24 ; Search time 71 Seconds

(without alignments)  
4047.048 Million cell updates/sec

Title: US-09-914-815A-10

Perfect score: 651  
Sequence: 1 ggaagaagacgcgcctggac.....cccgagcctaactctgcc 651

Scoring table:  
Gapop 10.0, Gapext 1.0

Searched: 56978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents NA:\*

- 1: /cgn2\_6/prodata/2/ina/5A\_COMB.seq:\*
- 2: /cgn2\_6/prodata/2/ina/5B\_COMB.seq:\*
- 3: /cgn2\_6/prodata/2/ina/6A\_COMB.seq:\*
- 4: /cgn2\_6/prodata/2/ina/6B\_COMB.seq:\*
- 5: /cgn2\_6/prodata/2/ina/PCTUS\_COMB.seq:\*
- 6: /cgn2\_6/prodata/2/ina/backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.2	14.3	642	4	US-09-370-838-119 Sequence 119, App
2	69.6	10.7	352	4	US-09-702-705-1204 Sequence 1204, App
3	69.6	10.7	352	4	US-09-736-457-1204 Sequence 1204, App
4	68	10.4	207	4	US-09-702-705-998 Sequence 998, App
5	68	10.4	207	4	US-09-736-457-998 Sequence 998, App
6	64.6	9.9	907	4	US-09-620-312D-652 Sequence 652, App
7	45.6	7.0	1311	4	US-09-252-991A-14546 Sequence 14546, A
8	45.6	7.0	3396	4	US-09-252-991A-14676 Sequence 14676, A
9	45.6	7.0	3444	4	US-09-252-991A-15078 Sequence 15078, A
10	45.2	6.9	510	4	US-09-252-991A-14818 Sequence 14818, A
11	43.2	6.6	745	1	US-08-036-555B-163 Sequence 163, App
12	43.2	6.6	745	1	US-08-469-569-163 Sequence 163, App
13	43.2	6.6	745	1	US-08-469-569-163 Sequence 163, App
14	43.2	6.6	745	1	US-08-249-322A-163 Sequence 163, App
15	43.2	6.6	745	1	US-08-469-526A-163 Sequence 163, App
16	43.2	6.6	745	2	US-08-734-591A-163 Sequence 163, App
17	43.2	6.6	745	2	US-08-469-660-163 Sequence 163, App
18	43.2	6.6	745	3	US-08-341-018-51 Sequence 51, Appl
19	43.2	6.6	745	3	US-08-470-335-163 Sequence 163, App
20	43.2	6.6	745	3	US-08-735-021-163 Sequence 163, App
21	43.2	6.6	745	3	US-08-734-664A-163 Sequence 163, App
22	43.2	6.6	745	4	US-08-470-339-163 Sequence 163, App
23	43.2	6.6	745	4	US-08-467-602-163 Sequence 163, App
24	43.2	6.6	745	5	PCT-US94-05083C-159 Sequence 159, App
25	43.2	6.6	745	5	PCT-US95-06846A-163 Sequence 163, App
26	43.2	6.6	2003	1	US-08-036-555B-21 Sequence 21, Appl
27	43.2	6.6	2003	1	US-08-469-569-21 Sequence 21, Appl

28	43.2	6.6	2003	1	US-08-469-526A-21 Sequence 21, Appl
29	43.2	6.6	2003	2	US-08-734-591A-21 Sequence 21, Appl
30	43.2	6.6	2003	2	US-08-469-660-21 Sequence 21, Appl
31	43.2	6.6	2003	3	US-08-341-018-71 Sequence 71, Appl
32	43.2	6.6	2003	3	US-08-470-335-21 Sequence 21, Appl
33	43.2	6.6	2003	3	US-08-735-021-21 Sequence 21, Appl
34	43.2	6.6	2003	3	US-08-734-664A-21 Sequence 21, Appl
35	43.2	6.6	2003	3	US-08-470-339-21 Sequence 21, Appl
36	43.2	6.6	2003	4	US-08-467-602-21 Sequence 21, Appl
37	43.2	6.6	2003	5	PCT-US94-05083C-21 Sequence 21, Appl
38	43.2	6.6	2003	5	PCT-US95-06846A-21 Sequence 21, Appl
39	41.2	6.3	764	3	US-08-991-789A-283 Sequence 283, App
40	41.2	6.3	764	4	US-09-062-451-283 Sequence 283, App
41	41.2	6.3	764	4	US-09-289-198-283 Sequence 283, App
42	40.6	6.2	2846	3	US-09-613-182-5 Sequence 5, Appl
43	39.8	6.1	2616	4	US-09-252-991A-16042 Sequence 16042, A
44	39.8	6.1	2739	4	US-09-252-991A-16558 Sequence 16558, A
45	39.4	6.1	615	4	US-09-501-115-21 Sequence 21, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-370-838-119
; Sequence 119, Application US/09370838
; Patent No. 6444425
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Mohamath, Roadoh
; APPLICANT: Sezarist, Heather
; TITLE OF INVENTION: COMPOUNDS FOR THERAPY AND DIAGNOSIS OF
; TITLE OF INVENTION: LUNG CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.47501
; CURRENT APPLICATION NUMBER: US/09/370, 838
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/285,323
; EARLIER FILING DATE: 1999-04-02
; NUMBER OF SEQ ID NOS: 289
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 119
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-370-838-119

Query Match      14.3%; Score 93.2; DB 4; Length 642;
Best Local Similarity 66.3%; Pred.No. 6.8e-13;
Matches 134; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

QY      125 GACCACTTATCTGCTGCTTCAAGCAACCTTCACTGATCTGTGCTCGGCTTC 184
      |||
DB      216 GACCATGTGTCTGTGCTTCAACACCTTCACTGATCTGTGCTCGGCTTC 275

QY      185 CTGGCGCTGCTTCACTCAAGCCGAGATCAGAACTGTGTGTAAGCTGAGCG 244
      |||
DB      276 ATAGCACTTGTCTTCACTCAAGCTGAGGAGGAGGAGATGTGCGACGACCGG 335

QY      245 GCCCGCGTTTGGCTTCCAAAGCTCAACATCTGCGCGGCGAGTGTGAGCGTGC 304
      |||
DB      336 GCCGAGGCTTATGCTTCAACCGGCAAGTCTGAGATCTGGGCTTGTGGGCACTC 395

QY      305 GTGCCGCACTGCTGCTCGG 326
      |||
DB      396 CTCATGACCATCTGTGTAATCG 417

RESULT 2
US-09-702-705-1204
; Sequence 1204, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
```

```
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1204
LENGTH: 352
TYPE: DNA
US-09-702-705-1204
ORGANISM: Homo sapien
```

```
Query Match 10.7%; Score 69.6; DB 4; Length 352;
Best Local Similarity 56.6%; Pred. No. 1.6e-07;
Matches 129; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
```

```
QY 23 ATGACACGCGGTATCCCGGAGACACCCGCGCCCGCAGCAGAGCCCGGTGCC 82
DB 123 ATGACACGAGGAGACATGAGGTGGCTGTGGGGGACCCCGCAGACCATCTTCCA 182
QY 83 CACACAGCCCTCACACTGGCGGCCCCCGCCCTCGAGACACTTATCTGTGTG 142
DB 183 AGGTCCACCGTATCAATCCACAGCAGACCTCGGCCGACCATGTCTGTGTGC 242
QY 143 GTGTTCAGACCCCTTACCTGATCTGTGTGCTCTGCGCTTCTGGCGCTGACTTCC 202
DB 243 CTGTTCACACCCCTTCTTCTGACATGCTGTCTGTGGCTTATGACTGCTTCTCC 302
QY 203 ATCAAGCCCGGATCAGAGGTGTGTGTGACTGTGAAGCCGCGCCGG 250
DB 303 GTGAAGTCTAGGAGACAGAAAGTGTGGCGACGTGACCGGGGCCGAG 350
```

```
RESULT 3
US-09-736-457-1204
Sequence 1204, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
```

```
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aijun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1204
LENGTH: 352
TYPE: DNA
US-09-736-457-1204
ORGANISM: Homo sapien
```

```
Query Match 10.7%; Score 69.6; DB 4; Length 352;
Best Local Similarity 56.6%; Pred. No. 1.6e-07;
```

```
Matches 129; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 23 ATGACACGCGGTATCCCGGAGACACCCGCGCCCGCAGCAGAGCCCGGTGCC 82
DB 123 ATGACACGAGGAGACATGAGGTGGCTGTGGGGGACCCCGCAGACCATCTTCCA 182
QY 83 CACACAGCCCTCACACTGGCGGCCCCCGCCCTCGAGACACTTATCTGTGTG 142
DB 183 AGGTCCACCGTATCAATCCACAGCAGACCTCGGCCGACCATGTGTGTGTGC 242
QY 143 GTGTTCAGACCCCTTACCTGATCTGTGTGCTCTGCGCTTCTGGCGCTGACTTCC 202
DB 243 CTGTTCACACCCCTTCTTCTGACATGCTGTGTGGCTTATGACTTGTGCTTCTCC 302
QY 203 ATCAAGCCCGGATCAGAGGTGTGTGTGACTGTGAAGCCGCGCCGG 250
DB 303 GTGAAGTCTAGGAGACAGAAAGTGTGGCGACGTGACCGGGGCCGAG 350
```

```
RESULT 4
US-09-702-705-998
Sequence 998, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
```

```
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darrick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 998
LENGTH: 207
TYPE: DNA
ORGANISM: Homo sapien
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(207)
OTHER INFORMATION: n = A,T,C or G
US-09-702-705-998
```

```
Query Match 10.4%; Score 68; DB 4; Length 207;
Best Local Similarity 71.8%; Pred. No. 3.4e-07;
Matches 89; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```
QY 125 GACCACTTATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 184
DB 83 GACCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 142
QY 185 CTGCGCGCGCTTACCTCATCAGAGCCGAGATCAGAGGTGTGTGTGTGTGTGTGT 244
DB 143 ATAGCATTCCTTCTTCTGAGAGTCTAGGACAGAAAGTGTGGCGAGCTGACGGG 202
QY 245 GCCC 248
DB 203 GCCC 206
```

```
RESULT 5
US-09-736-457-998
Sequence 998, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
```

```
APPLICANT: Wang, Tongtong
```







```
RESULT 11
US-08-036-555B-163
; Sequence 163, Application US/08036555B
; Patent No. 5530109
; GENERAL INFORMATION:
; APPLICANT: Goodheart, Andrew; Stroobant, Paul;
; APPLICANT: Minshetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Miao Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/036,555B
; FILING DATE: 24-MAR-1993
; CLASSIFICATION: 435
; APPLICATION DATA:
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-036-555B-163

Query Match 6.6%; Score 43.2; DB 1; Length 745;
Best Local Similarity 57.4%; Pred. No. 0.21;
Matches 78; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 246 CCGGCGCTTTGGCTCCAAAGCCCAAGTGTACAAATCTCGGCGCGATGTGACGCTGG 305
DB 29 CCGGCGCTCCCGGCGCCCGGCGCCAGCGCCCGGCTCGCGCGCTCGTCCGCGCGC 88
QY 306 TCGCGCACGTGCTGCTCTTGAGGCTGTGTGATGTGCTGACCTGAGCCCGGCTGG 365
DB 89 TCGCGCTGCTGCGCACTGCTGCTGTGAGGACCGCGGCTTGGCGCGGAGCGCGG 148
QY 366 CCAAGACTGTGCGCGC 381
```

```
DB 149 CCGGCAACGAGCGCGC 164

RESULT 12
US-08-469-569-163
; Sequence 163, Application US/08469569
; Patent No. 5606032
; GENERAL INFORMATION:
; APPLICANT: Goodheart, Andrew; Stroobant, Paul;
; APPLICANT: Minshetti, Luisa; Waterfield, Michael; Marchioni, Mark;
; APPLICANT: Chen, Miao Su; Hiles, Ian
; TITLE OF INVENTION: Glial Mitogenic Factors, Their
; TITLE OF INVENTION: Preparation and Use
; NUMBER OF SEQUENCES: 184
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,569
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; APPLICATION DATA:
; APPLICATION NUMBER: 08/036,555
; FILING DATE: 24-MAR-1993
; APPLICATION NUMBER: 07/965,173
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/940,389
; FILING DATE: 03-SEP-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/907,138
; FILING DATE: 30-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/863,703
; FILING DATE: 03-APRIL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: U.K. 91 07566.3
; FILING DATE: 10-APRIL-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Tsai, Christine H.
; REGISTRATION NUMBER: 34,266
; REFERENCE/DOCKET NUMBER: LUD 5250.4
; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 163:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-469-569-163

Query Match 6.6%; Score 43.2; DB 1; Length 745;
Best Local Similarity 57.4%; Pred. No. 0.21;
Matches 78; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 246 CCGGCGCTTTGGCTCCAAAGCCCAAGTGTACAAATCTCGGCGCGATGTGACGCTGG 305
DB 29 CCGGCGCTCCCGGCGCCCGGCGCCAGCGCCCGGCTCGCGCGCTCGTCCGCGCGC 88
QY 306 TCGCGCACGTGCTGCTCTTGAGGCTGTGTGATGTGCTGACCTGAGCCCGGCTGG 365
DB 89 TCGCGCTGCTGCGCACTGCTGCTGTGAGGACCGCGGCTTGGCGCGGAGCGCGG 148
```

Qy 366 CCAAGACTCTGCGC 381  
Db 149 CCGGCAACGAGCGGC 164

## RESULT 13

US-08-249-322A-163  
Sequence 163, Application US/08249322A  
Patent No. 5716930

## GENERAL INFORMATION:

APPLICANT: Goodearl, Andrew; Stroobant, Paul;  
Minghetti, Luisa; Waterfield, Michael; Marchionni, Mark;  
APPLICANT: Chen, Maio Su; Hiles, Ian  
TITLE OF INVENTION: Glial Mitogenic Factors, Their  
NUMBER OF SEQUENCES: 184  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pelite & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/249,322A  
FILING DATE: 26-MAY-1994

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APRIL-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Tsai, Christine H.  
REGISTRATION NUMBER: 34,266

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884

INFORMATION FOR SEQ ID NO: 163:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 745  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-249-322A-163

Query Match  
Best Local Similarity 57.4%; Score 43.2; DB 1; Length 745;  
Matches 78; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 246 CCGGCGTTTGGCTCCAAAGCCCAAGTGTACAAATCTGGCCGCGATGTGAGCGCTGG 305  
Db 29 CCGGCGTCCCGGCCCCCGGCGCCAGCGCCCGGCTCCGCGCGCTGCTGTGCGCGCGC 88

Qy 306 TGCCGCACTGCTGCTCTGCGGCGCTGCTGTAAGTGTGCTGCACTTGGCCCCGCGCTGG 365  
Db 89 TGCCGCTGCTGCACTGCTGCTGCGGCAAGCGGCGCTGCGCGCGCGCGCG 148  
Qy 366 CCAAGACTCTGCGC 381  
Db 149 CCGGCAACGAGCGGC 164

## RESULT 14

US-08-469-526A-163  
Sequence 163, Application US/08469526A  
Patent No. 5792849

## GENERAL INFORMATION:

APPLICANT: Goodearl, Andrew  
APPLICANT: Stroobant, Paul  
APPLICANT: Minghetti, Luisa  
APPLICANT: Waterfield, Michael  
APPLICANT: Marchionni, Mark  
APPLICANT: Chen, Maio Su  
APPLICANT: Hiles, Ian  
TITLE OF INVENTION: GLIAL MITOGENIC FACTORS, THEIR  
NUMBER OF SEQUENCES: 187  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Clark & Biring LLP  
STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: PastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,526A  
FILING DATE: 06 June 1995

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/036,555  
FILING DATE: 24-MAR-1993

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/965,173  
FILING DATE: 23-OCT-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/940,389  
FILING DATE: 03-SEP-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/907,138  
FILING DATE: 30-JUN-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/863,703  
FILING DATE: 03-APRIL-1992

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: U.K. 91 07566.3  
FILING DATE: 10-APR-1991

ATTORNEY/AGENT INFORMATION:  
NAME: Bleker-Brady, Kristina  
REGISTRATION NUMBER: 39,109

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045

INFORMATION FOR SEQ ID NO: 163:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 745  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-469-526A-163

Query Match  
Best Local Similarity 57.4%; Score 43.2; DB 1; Length 745;  
Matches 78; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 246 CCGGCGTTTGGCTCCAAAGCCCAAGTGTACAAATCTGGCCGCGATGTGAGCGCTGG 305  
Db 29 CCGGCGTCCCGGCCCCCGGCGCCAGCGCCCGGCTCCGCGCGCTGCTGTGCGCGCGC 88



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 12:18:23 ; Search time 286 Seconds  
(without alignments)  
6144.525 Million cell updates/sec

Title: US-09-914-815A-10  
Perfect score: 1 ggaagagacggcgtggaac.....ccgggacctactctgcc 651  
Sequence: 651

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	651	100.0	651	21	AAA64687
2	424.6	65.2	651	24	AB189793
3	96.4	14.8	447	22	AA119241
4	95.4	14.7	326	22	ABA47169
5	95.4	14.7	326	22	ABA65053
6	95.4	14.7	326	22	ABA32159
7	95.4	14.7	326	22	AAK13477
8	95.4	14.7	326	22	AAK39215

C	9	95.4	14.7	326	22	AA170025	Probe #958 for ge
C	10	95.4	14.7	326	22	AA145223	Probe #13909 used
C	11	95.4	14.7	326	22	AA105732	Probe #5723 used t
C	12	95.4	14.7	326	23	AB388800	Human liver single
C	13	95.4	14.7	326	24	AB313299	Human genome-deriv
C	14	95.4	14.7	459	22	ABA42012	Human breast cell
C	15	95.4	14.7	459	22	ABA52433	Human foetal liver
C	16	95.4	14.7	459	22	ABA22426	Probe #692 for gen
C	17	95.4	14.7	459	22	AAK00702	Human brain expres
C	18	95.4	14.7	459	22	AAK6153	Human bone marrow
C	19	95.4	14.7	459	22	AA110783	Probe #716 for gen
C	20	95.4	14.7	459	22	AA112041	Probe #727 used to
C	21	95.4	14.7	459	22	AA100708	Probe #699 used to
C	22	95.4	14.7	459	23	AB525742	Human liver single
C	23	95.4	14.7	459	24	AB500738	Human genome-deriv
C	24	95.4	14.6	407	22	AA126291	Human breast cance
C	25	93.4	14.3	621	25	ACC42315	Human MAP kinase c
C	26	93.4	14.3	631	24	AB099282	Human coding beque
C	27	93.4	14.3	905	24	ABK84587	Human cDNA differe
C	28	93.4	14.3	905	24	ABK69222	Prostate cancer re
C	29	93.4	14.3	905	25	ACC42316	Human MAP kinase c
C	30	93.4	14.3	905	25	ACC42354	Human MAP kinase c
C	31	93.4	14.3	905	25	ACC42355	Human MAP kinase c
C	32	93.2	14.3	484	24	AB060102	Human colon cancer
C	33	93.2	14.3	606	25	ACC42317	Human MAP kinase c
C	34	93.2	14.3	621	25	ACC42319	Human MAP kinase c
C	35	93.2	14.3	642	20	AA207228	Human lung tumour
C	36	93.2	14.3	642	21	AA279193	Human lung tumour
C	37	93.2	14.3	642	21	AA231193	Human lung tumour
C	38	93.2	14.3	695	24	AA595009	Human DNA sequence
C	39	93.2	14.3	732	23	AA575071	DNA encoding novel
C	40	93.2	14.3	761	24	AB056182	Human ovarian anti
C	41	93.2	14.3	808	25	AAA46668	CDNA of I-BU gene
C	42	93.2	14.3	808	25	ACC42318	Human MAP kinase c
C	43	92.4	14.2	637	23	AA575776	DNA encoding novel
C	44	91.8	14.1	716	22	AA159665	Human polynucleoti
C	45	91.8	14.1	999	22	AA157879	Human polynucleoti

ALIGNMENTS

RESULT 1	AAA64687	standard; cDNA; 651 BP.
ID	AAA64687	
AC	AAA64687	
XX	02-JAN-2001	(first entry)
XX		
DE		cDNA encoding a human leukocyte and blood related protein (LBAP).
XX		
XX		Human: leukocyte and blood related protein; LBAP; arteriosclerosis;
KW		cell proliferative disorder; actinic keratosis; atherosclerosis;
KW		burialitis; cirrhosis; hepatitis; mixed connective tissue disease; MCTD;
KW		myelofibrosis; paroxysmal nocturnal hemoglobinuria; cancer;
KW		adenocarcinoma; leukemia; lymphoma; melanoma; myeloma; sarcoma;
KW		teratocarcinoma; autoimmune disorder; inflammatory disorder;
KW		acquired immunodeficiency syndrome; AIDS; Addison's disease;
KW		adult respiratory distress syndrome; allergy; ankylosing spondylitis;
KW		amyloidosis; anaemia; asthma; autoimmune haemolytic anaemia; infection;
KW		Werner syndrome; haemodialysis; extracorporeal circulation; trauma; ss.
XX		
OS		Homo sapiens.
XX		
FT	Key	Location/Qualifiers
FT	CDS	23..421
FT		/tag= a
XX		/product= "leukocyte and blood related protein (LBAP)"
XX		
PN		MO200052161-A2.
XX		
PD		08-SEP-2000.



XX SQ Sequence 451 BP; 74 A; 156 C; 137 G; 83 T; 1 other;  
Query Match 65.2%; Score 424.6; DB 24; Length 451;  
Best Local Similarity 98.4%; Pred. No. 1.1e-84;  
Matches 438; Conservative 1; Mismatches 5; Indels 1; Gaps 1;  
OY 7 GACGGCGCTGAGACCATGAGACAGCGGCTATCCCGGAGACACCCGGGCTCCAGAGCC 66  
DB 6 GAGGGCGCTGAGACCATGAGACAGCGGCTATCCCGGAGACACCCGGGCTCCAGAGCC 65  
OY 67 CAGCAAGCCCGGTGGCCACACAGCCCTGACACTGAGCGGCCCGACCCCGGCTCGAGA 126  
DB 66 CAGCAAGCCCGGTGGCCACACAGCCCTGACACTGAGCGGCCCGACCCCGGCTCGAGA 125  
OY 127 CCACTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 186  
DB 126 CCACTTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 185  
OY 187 GACGCTGAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 246  
DB 186 GACGCTGAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245  
OY 247 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 306  
DB 246 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305  
OY 307 GCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 366  
DB 306 GCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 364  
OY 367 CAAGAGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 426  
DB 365 CAAGAGCTGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 424  
OY 427 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 451  
DB 425 TGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 449

RESULT 3  
AAL19241/c  
ID AAL19241 standard; cDNA; 447 BP.  
XX  
AC AAL19241;  
XX  
DT 07-DEC-2001 (first entry)  
XX  
DE Human breast cancer expressed polynucleotide 11698.  
XX  
KW Human; breast cancer; cell marker; cytosolic; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200151628-A2.  
XX  
PD 19-JUL-2001.  
XX  
PF 10-JAN-2001; 2001MO-US00798.  
XX  
PR 14-JAN-2000; 2000US-0176077.  
XX  
PR 24-MAR-2000; 2000US-0189167.  
XX  
PR 22-MAR-2000; 2000US-0192089.  
XX  
PR 15-MAY-2000; 2000US-0193480.  
XX  
PR 09-JUN-2000; 2000US-0205230.  
XX  
PR 25-JUL-2000; 2000US-0211315.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Lillie J, Xu Y, Wang Y, Steinmann K;  
XX  
DR WPI; 2001-451856/48.

XX FT New peptide useful as a marker for the diagnosis of breast cancer -  
XX PS Claim 1; Page 2084; 3695pp; English.  
XX  
XX The invention relates to human breast cancer expressed polynucleotides  
CC (AAL07544-AAL26789) and methods of assessing whether a patient is  
CC afflicted with breast cancer by examining the correlation between the  
CC expression of certain markers and the cancerous state of breast cells.  
CC The polynucleotides and encoded polypeptides are potential markers for  
CC detecting, diagnosing, monitoring, characterizing treating and  
CC potentially preventing breast cancer. The polynucleotides and encoded  
CC polypeptides are also useful for isolating compounds with cytostatic  
CC activity.  
XX  
SQ Sequence 447 BP; 102 A; 101 C; 121 G; 123 T; 0 other;  
Query Match 14.8%; Score 96.4; DB 22; Length 447;  
Best Local Similarity 62.4%; Pred. No. 3.6e-12;  
Matches 151; Conservative 0; Mismatches 91; Indels 0; Gaps 0;  
OY 114 CCCGCTCGAGACCACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 173  
DB 421 CCGTGGCTGAGACCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 362  
OY 174 GCTTGGCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 233  
DB 361 GCTTGGCTTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 302  
OY 234 ACCGGAAGGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 293  
DB 301 ACCGGAAGGGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 242  
OY 294 TGTGAGCGTGTGCGCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 353  
DB 241 TTTTGAACATCTTCATGACCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 182  
OY 354 TG 355  
DB 181 AG 180

RESULT 4  
ABA47169/c  
ID ABA47169 standard; DNA; 326 BP.  
XX  
AC ABA47169;  
XX  
DT 01-FEB-2002 (first entry)  
XX  
DE Human breast cell single exon nucleic acid probe #5864.  
XX  
KW Human; microarray; single exon probe; gene expression; breast;  
XX  
KW disease; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157271-A2.  
XX  
PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001MO-US00662.  
XX  
PR 04-FEB-2000; 2000US-0180312.  
XX  
PR 26-MAY-2000; 2000US-0207456.  
XX  
PR 30-JUN-2000; 2000US-0608408.  
XX  
PR 03-AUG-2000; 2000US-0632366.  
XX  
PR 21-SEP-2000; 2000US-0234687.  
XX  
PR 27-SEP-2000; 2000US-0236359.  
XX  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLB-) MOLECULAR DYNAMICS INC.  
XX  
PI



PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-49693/54.  
XX  
XX  
PT New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
PT breast, comprises number of single exon nucleic acid probes -  
XX  
XX  
PS Claim 4; SEQ ID NO 5864; 327bp + sequence listing; English.  
XX  
XX  
CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human breast and BT 474 cells. The method involves contacting  
CC the probes with a collection of detectably labelled nucleic acids  
CC derived from mRNA of human breast, and then measuring the label  
CC bound to each probe of the microarray. The probes are useful for  
CC verifying the expression of regions of genomic DNA predicted to  
CC encode proteins. They are useful for gene discovery, and for  
CC determining predisposition and/or prognosing breast disease. Gene  
CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTGATGCTGGTGGTTCAGACCCCTTACTGATGTTGTTGCTGCTCC 185  
DB 209 ACCATGTTGTGTCCTGCTTCAACACCTCTTATGAACCCCTGCTGCTGCTTCA 150  
QY 186 TGGCGTGGCTACTCTCATCAAGCCCGAGATCAAGAGTGTGTGCTGGAAGCG 245  
DB 149 TAGGCTTACCTACTCATCAAGTCTAGGAGACAGAAAGTGTGGCACTGACCGGG 90  
QY 246 CCCGCGTTTGGCTCCAAAGCCCAAGTCTTACAACATCTGCGCGATGTGAAGCTGG 305  
DB 89 CCCAGGCTATGCTCTCAACCGCCAAAGTCTGAACATCTGGGCTTGAATTGGGCAATCC 30  
QY 306 TGGCGCACTGCTGCTCTCTGG 326  
DB 29 TCATGACCATCTCTGCTCATCG 9  
RESULT 5  
ABA65053/C  
ID ABA65053 standard; DNA; 326 BP.  
XX  
XX  
AC ABA65053;  
XX  
XX  
DT 01-FEB-2002 (first entry)  
XX  
XX  
DE Human foetal liver single exon nucleic acid probe #13358.  
XX  
XX  
KW Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200157277-A2.  
XX  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 30-JAN-2001; 2001WO-US00669.  
XX  
XX  
PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX  
DR WPI; 2001-48344/52.  
XX  
XX  
PT Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human fetal liver -  
XX  
XX  
PS Claim 4; SEQ ID NO 13358; 639bp + sequence listing; English.  
XX  
XX  
CC The invention relates to a single exon nucleic acid probe for  
CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC fetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTGATGCTGGTGGTTCAGACCCCTTACTGATGTTGTTGCTGCTCC 185  
DB 209 ACCATGTTGTGTCCTGCTTCAACACCTCTTATGAACCCCTGCTGCTGCTTCA 150  
QY 186 TGGCGTGGCTACTCTCATCAAGCCCGAGATCAAGAGTGTGTGCTGGAAGCG 245  
DB 149 TAGGCTTACCTACTCATCAAGTCTAGGAGACAGAAAGTGTGGCACTGACCGGG 90  
QY 246 CCCGCGTTTGGCTCCAAAGCCCAAGTCTTACAACATCTGCGCGATGTGAAGCTGG 305  
DB 89 CCCAGGCTATGCTCTCAACCGCCAAAGTCTGAACATCTGGGCTTGAATTGGGCAATCC 30  
QY 306 TGGCGCACTGCTGCTCTCTGG 326  
DB 29 TCATGACCATCTCTGCTCATCG 9  
RESULT 6  
ABA32159/C  
ID ABA32159 standard; DNA; 326 BP.  
XX  
XX  
AC ABA32159;  
XX  
XX  
DT 23-JAN-2002 (first entry)  
XX  
XX  
DE Probe #10625 for gene expression analysis in human heart cell sample.  
XX  
XX  
KW Human; gene expression; heart; microarray; vascular system; probe;  
KW cardiovascular disease; hypertension; cardiac arrhythmia;  
KW congenital heart disease; ss.  
XX  
XX  
OS Homo sapiens.  
XX  
XX  
PN WO200157274-A2.  
XX  
XX  
PD 09-AUG-2001.  
XX  
XX  
PF 30-JAN-2001; 2001WO-US00666.  
XX

PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-488699/53.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX hearts -  
XX  
PS Claim 4; SEQ ID No 10625; 530pp; English.  
XX  
CC The present invention relates to single exon nucleic acid probes for  
CC measuring human gene expression in a sample derived from human heart. The  
CC present sequence is one such probe. The probes may be used for  
CC predicting, measuring and displaying gene expression in samples derived  
CC from the human heart via microarrays. By measuring gene expression, the  
CC probes are useful for predicting, diagnosing, grading, staging,  
CC monitoring and prognosing diseases of the human heart and vascular system  
CC e.g. cardiovascular disease, hypertension, cardiac arrhythmias and  
CC congenital heart disease.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_ptc\_sequences.  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
XX  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
XX  
QY 126 ACCACTGATCTGTGCTGGTTCAGACCCCTTCACTGTAATCTGTGCTCGCTTCC 185  
Db 209 ACATGTTGTCTGTGCTGCTTCAACCTCTTCAATGAACCCCTGCTGCTGCTTCA 150  
QY 186 TGGGCTGGCTTACTCATCAAGCCCGAATCAAGAGTGTGTGATCTGGAACGG 245  
Db 149 TAGGTTCACTTACTCATGAAGCTTAGGGAAGAAGATGTGTGCGACCTGACCGGG 90  
QY 246 CCCGCGCTTTGGCTCCAAAGCAAGTGTACCAATCTGCGCGCATGTGACGCTGG 305  
Db 89 CCCAGGCTTATGCTTCCACCGCCAAAGTGTCTGAACATCTGGGCCCTTGACTTTGGGCATCC 30  
QY 306 TGCCGCCACTGTGCTGCTCTGG 326  
Db 29 TCATGACCATTTGCTCATCG 9  
XX  
RESULT 7  
AAK13477/c  
ID AAK13477 standard; DNA; 326 BP.  
XX  
AC AAK13477;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human brain expressed single exon probe SEQ ID NO: 13468.  
XX  
KW Human; brain expressed exon; gene expression analysis; probe;  
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;  
KW epilepsy; cancer; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157275-A2.  
XX

PD 09-AUG-2001.  
XX  
PF 30-JAN-2001; 2001WO-US00667.  
XX  
XX  
PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
DR WPI; 2001-483446/52.  
XX  
PT Single exon nucleic acid probes for analyzing gene expression in human  
XX brains -  
XX  
PS Example 4; SEQ ID NO: 13468; 650pp + Sequence Listing; English.  
XX  
CC The present invention provides a number of single exon nucleic acid  
CC probes which are derived from genomic sequences expressed in the human  
CC brain. They can be used to measure gene expression in brain cell samples,  
CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is one of the probes of the  
CC invention.  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
XX  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
XX  
QY 126 ACCACTGATCTGTGCTGGTTCAGACCCCTTCACTGTAATCTGTGCTCGCTTCC 185  
Db 209 ACATGTTGTCTGTGCTGCTTCAACCTCTTCAATGAACCCCTGCTGCTGCTTCA 150  
QY 186 TGGGCTGGCTTACTCATCAAGCCCGAATCAAGAGTGTGTGATCTGGAACGG 245  
Db 149 TAGGTTCACTTACTCATGAAGCTTAGGGAAGAAGATGTGTGCGACCTGACCGGG 90  
QY 246 CCCGCGCTTTGGCTCCAAAGCAAGTGTACCAATCTGCGCGCATGTGACGCTGG 305  
Db 89 CCCAGGCTTATGCTTCCACCGCCAAAGTGTCTGAACATCTGGGCCCTTGACTTTGGGCATCC 30  
QY 306 TGCCGCCACTGTGCTGCTCTGG 326  
Db 29 TCATGACCATTTGCTCATCG 9  
XX  
RESULT 8  
AAK39215/c  
ID AAK39215 standard; DNA; 326 BP.  
XX  
AC AAK39215;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human bone marrow expressed single exon probe SEQ ID NO: 13772.  
XX  
KW Human; bone marrow expressed exon; gene expression analysis; probe;  
KW microarray; cancer; leukemia; lymphoma; myeloma; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200157276-A2.  
XX  
PD 09-AUG-2001.  
XX



PR 04-FEB-2000; 2000US-0180312.  
PR 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024253.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-488897/53.  
XX  
XX Human genome-derived single exon nucleic acid probes useful for  
XX analyzing gene expression in human placenta -  
XX  
XX Claim 25; SEQ ID No 13909; 654bp; English.  
XX  
XX The present invention relates to single exon nucleic acid probes (SENP).  
XX The present sequence is one such probe. The probes are useful for  
XX producing a microarray for predicting, measuring and displaying gene  
XX expression in samples derived from human placenta. The probes are useful  
XX for antenatal diagnosis of human genetic disorders.  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTTGAATCTGCTGCTGCTTCAAGACCCCTTCACTGAATCTGCTGCTGCTTCC 185  
DB 209 ACCATGTCTGCTGCTGCTGCTTCAAGACCCCTTCACTGAATCTGCTGCTGCTTCA 150  
QY 186 TGGCGTGGCTCTCATCAAGCCCGAGATCAAGAGTGTTGGTGAACCTGGAAGCGG 245  
DB 149 TAGCGTTCACTTCTCATGAAGCTTAAGGACAGAAAGATGTTGGACCTTAACCGGG 90  
QY 246 CCGGCGGTTTGGCTCCAAAGCCAAAGTGTACAACTCTGCGCGCATGTGACGCTGG 305  
DB 89 CCCAGGCGTATGCTCCACCGCCCAAGTGCCTGAACATCTGGGCCCTGACTTTGGGATCC 30  
QY 306 TGGCGCACTGCTGCTCTCTGG 326  
DB 29 TCATGACCAATCTGCTCATCG 9  
RESULT 11  
AAI05732/C  
ID AAI05732 standard; DNA; 326 BP.  
XX  
XX AAI05732;  
XX  
XX 09-OCT-2001 (first entry)  
XX  
XX Probe #5723 used to measure gene expression in human breast sample.  
XX  
XX KM Inflammatory disease; breast cancer; development disorder; ss;  
XX inflammatory disease; proliferative breast disease; non-carcinoma tumour.  
XX  
XX Homo sapiens.  
XX  
XX MO200157270-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX  
XX 29-JAN-2001; 2001WO-US00661.  
XX  
XX 04-FEB-2000; 2000US-0180312.  
XX  
XX 26-MAY-2000; 2000US-0207456.  
XX  
XX 30-JUN-2000; 2000US-0608408.  
XX  
XX 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
PR 04-OCT-2000; 2000GB-0024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
XX WPI; 2001-476286/51.  
XX  
XX Novel single exon nucleic acid probe used to measuring gene expression  
XX in a human breast -  
XX  
XX Claim 25; SEQ ID No 5723; 322bp; English.  
XX  
XX The present invention relates to novel single exon nucleic acid probes.  
XX The present sequence is one such probe. The probes are useful for  
XX measuring human gene expression in a human breast sample, where the probe  
XX hybridizes at high stringency to a nucleic acid expressed in the human  
XX breast. The probes are useful for predicting, diagnosing, grading,  
XX staging, monitoring and prognosing diseases of the human breast,  
XX particularly those diseases with polygenic aetiology. The diseases  
XX include: breast cancer, disorders of development, inflammatory diseases  
XX of the breast, fibrocystic changes, proliferative breast disease and  
XX non-carcinoma tumours.  
XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 22; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTTGAATCTGCTGCTGCTTCAAGACCCCTTCACTGAATCTGCTGCTGCTTCC 185  
DB 209 ACCATGTCTGCTGCTGCTGCTTCAAGACCCCTTCACTGAATCTGCTGCTGCTTCA 150  
QY 186 TGGCGTGGCTCTCATCAAGCCCGAGATCAAGAGTGTTGGTGAACCTGGAAGCGG 245  
DB 149 TAGCGTTCACTTCTCATGAAGCTTAAGGACAGAAAGATGTTGGACCTTAACCGGG 90  
QY 246 CCGGCGGTTTGGCTCCAAAGCCAAAGTGTACAACTCTGCGCGCATGTGACGCTGG 305  
DB 89 CCCAGGCGTATGCTCCACCGCCCAAGTGCCTGAACATCTGGGCCCTGACTTTGGGATCC 30  
QY 306 TGGCGCACTGCTGCTCTCTGG 326  
DB 29 TCATGACCAATCTGCTCATCG 9  
RESULT 12  
ABS38800/C  
ID ABS38800 standard; DNA; 326 BP.  
XX  
XX ABS38800;  
XX  
XX 25-FEB-2003 (first entry)  
XX  
XX Human liver single exon probe, SEQ ID No 13790.  
XX  
XX KM Human; single exon nucleic acid probe; liver; cirrhosis;  
XX hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;  
XX coronary heart disease; ss.  
XX  
XX Homo sapiens.  
XX  
XX MO200157273-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 30-JAN-2001; 2001WO-US00664.



50 Sequence 326 BP; 70 A; 79 C; 113 G; 64 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 24; Length 326;  
Best Local Similarity 67.2%; Pred. No. 5.8e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTGATCTGGTGGTGTTCAGACCCCTCTGCTGAATCTGTGTGCTGCTGCTTCC 185  
DB 209 ACCATGTTCTGGTCCCTGTTCAACCCCTTTCATGAACCCCTGCTGCTGCGCTTCA 150  
QY 186 TGGGCGTGGCCCTACTCTCAATCAAGCCGAGATCAAGATGTTGTGTGACCTGGAAGCGG 245  
DB 149 TAGGTTACCTACTCCATGAAGTCTAGGACAGGAAGATGTTGGCGACCTGACCGCGG 90  
QY 246 CCCGCGTTTGGCTCCCAAGCCAAAGTGTACCAATCTCTGCGCGCATGTGAACGCTGG 305  
DB 89 CCCAGGCTATGCTCTCCACCGCCAAAGTCCCTGAACATCTGGGCTTGAATTTGGGCAATCC 30  
QY 306 TGGCGCCACTGCTGCTCTGG 326  
DB 29 TCATGACCATCTGCTCATCG 9  
RESULT 14  
ABA42012/C  
ID ABA42012 standard; DNA; 459 BP.  
AC ABA42012;  
XX ABA42012;  
XX 01-FEB-2002 (first entry)  
DT Human breast cell single exon nucleic acid probe #707.  
DE Human breast cell single exon nucleic acid probe #707.  
XX Human; microarray; single exon probe; gene expression; breast;  
KM disease; cancer; ss.  
XX Homo sapiens.  
OS  
XX WO200157271-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US00662.  
PF 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-496933/54.  
DR WPI; 2001-496933/54.  
XX New spatially-addressable set of single exon nucleic acid probes,  
PT useful for measuring gene expression in sample derived from human  
breast, comprises number of single exon nucleic acid probes  
XX Claim 1; SEQ ID NO 707; 327bp + sequence listing; English.  
XX The invention relates to a spatially-addressable set of single exon  
XX nucleic acid probes for measuring gene expression in a sample derived  
XX from human breast and BT 474 cells. The method involves contacting  
XX the probes with a collection of detectably labelled nucleic acids  
XX derived from mRNA of human breast, and then measuring the label  
XX bound to each probe of the microarray. The probes are useful for  
XX verifying the expression of regions of genomic DNA predicted to  
XX encode proteins they are useful for gene discovery, and for  
XX determining predisposition and/or prognosing breast disease. Gene

CC expression analysis is useful for assessing the toxicity of chemical  
CC agents on cells. The microarray of this invention presents a far greater  
CC diversity of probes for measuring gene expression, with far less bias  
CC than expressed sequence tag microarrays. The method is suitable for  
CC rapid production of functional information from genomic sequence. The  
CC present sequence is a single exon nucleic acid probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at [http://wipo.int/pub/published\\_pot\\_sequences](http://wipo.int/pub/published_pot_sequences).  
XX  
50 Sequence 459 BP; 115 A; 96 C; 154 G; 94 T; 0 other;  
Query Match 14.7%; Score 95.4; DB 22; Length 459;  
Best Local Similarity 67.2%; Pred. No. 6.1e-12;  
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;  
QY 126 ACCACTGATCTGGTGGTGTTCAGACCCCTCTGCTGAATCTGTGTGCTGCTGCTTCC 185  
DB 406 ACCATGTTCTGGTCCCTGTTCAACCCCTTTCATGAACCCCTGCTGCTGCGCTTCA 347  
QY 186 TGGGCGTGGCCCTACTCTCAATCAAGCCGAGATCAAGATGTTGTGTGACCTGGAAGCGG 245  
DB 346 TAGGTTACCTACTCCATGAAGTCTAGGACAGGAAGATGTTGGCGACCTGACCGCGG 287  
QY 246 CCCGCGTTTGGCTCCCAAGCCAAAGTGTACCAATCTCTGCGCGCATGTGAACGCTGG 305  
DB 286 CCCAGGCTATGCTCTCCACCGCCAAAGTCCCTGAACATCTGGGCTTGAATTTGGGCAATCC 227  
QY 306 TGGCGCCACTGCTGCTCTGG 326  
DB 226 TCATGACCATCTGCTCATCG 206  
RESULT 15  
ABA52433/C  
ID ABA52433 standard; DNA; 459 BP.  
AC ABA52433;  
XX ABA52433;  
XX 01-FEB-2002 (first entry)  
DT Human foetal liver single exon nucleic acid probe #738.  
DE Human foetal liver single exon nucleic acid probe #738.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe; ss.  
XX Homo sapiens.  
OS  
XX WO200157277-A2.  
PN 09-AUG-2001.  
PD 30-JAN-2001; 2001WO-US00669.  
PF 04-FEB-2000; 2000US-0180312.  
XX 26-MAY-2000; 2000US-0207456.  
PR 30-JUN-2000; 2000US-0608408.  
PR 03-AUG-2000; 2000US-0632366.  
XX 21-SEP-2000; 2000US-0234687.  
PR 27-SEP-2000; 2000US-0236359.  
XX 04-OCT-2000; 2000GB-0024263.  
PA (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-493447/52.  
DR WPI; 2001-493447/52.  
XX Human genome-derived single exon nucleic acid probes useful for  
PT analyzing gene expression in human foetal liver -  
XX Claim 1; SEQ ID NO 738; 639bp + sequence listing; English.  
XX The invention relates to a single exon nucleic acid probe for

CC measuring human gene expression in a sample derived from human foetal  
CC liver. The single exon nucleic acid probes may be used for predicting,  
CC measuring and displaying gene expression in samples derived from human  
CC foetal liver. The present sequence is a single exon nucleic acid  
CC probe of the invention.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

Sequence 459 BP; 115 A; 96 C; 154 G; 94 T; 0 other;

Query Match 14.7%; Score 95.4; DB 22; Length 459;  
Best Local Similarity 67.2%; Freq. No. 6.1e-12;

Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY	126	ACCACTTGATCTGGTGGGTTGTCAGACACCTCTACCTGAATCTGTGCTCGGCTTCC	185
Db	406	ACCATGTTGTGTGGTCCCTGTTCAACACCTCTTCATGAAACCTGCTGCGCTTCA	347
QY	186	TGGGCTGGCCCTACTCCATCAAGGCCGAGATCAGAAAGTGTGTGACCTGGAAGCGG	245
Db	346	TAGCGTTCACTACTCCATGAAGTCTAGGAGACAGGAAGATGTTGGCGACCTGACCGGGG	287
QY	246	CCCGGCGTTTGGCTCCAAAGCTGACATCTGCGCCGAGTGTGAAGCTGG	305
Db	286	CCCAAGGCTATGCTCCACCGCAAGTGTGAACATCTGGGCCCTGACTTTGGGCAATCC	227
QY	306	TGCCGCCACCTGCTGCTCTGG	326
Db	226	TCATGACCATTCTGCTCATCG	206

Search completed: January 31, 2004, 13:43:02

Job time : 290 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 31, 2004, 13:23:10 ; Search time 2842 Seconds

(without alignments)  
9370.924 Million cell updates/sec

Title: US-09-914-815A-10

Perfect score: 651  
Sequence: 1 ggaagagagcgcgcgcgaac.....cccgagccctaactctgcc 651

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBankl: \*  
1: gb\_ba: \*  
2: gb\_hcg: \*  
3: gb\_in: \*  
4: gb\_om: \*  
5: gb\_ov: \*  
6: gb\_pac: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sfb: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_pat: \*  
24: em\_ph: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sfb: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_hcg\_hum: \*  
31: em\_hcg\_inv: \*  
32: em\_hcg\_other: \*  
33: em\_hcg\_mus: \*  
34: em\_hcg\_pln: \*  
35: em\_hcg\_rod: \*  
36: em\_hcg\_mam: \*  
37: em\_hcg\_vtl: \*  
38: em\_sy: \*  
39: em\_hcg\_hum: \*  
40: em\_hcg\_mus: \*  
41: em\_hcg\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	446.6	68.6	95038	9 AF015416	AF015416 Homo sapi
C 2	446.6	68.6	143835	9 AC136475	AC136475 Homo sapi
C 3	446.6	68.6	156819	2 AP006283	AP006283 Homo sapi
C 4	446.6	68.6	175416	2 AP006286	AP006286 Homo sapi
C 5	446.6	68.6	187160	9 AC138230	AC138230 Homo sapi
C 6	284.6	43.7	499	10 MMU9781	MMU9781 Mus muscu
C 7	173	26.6	238533	2 AC132968	AC132968 Rattus no
C 8	167.6	25.7	200734	2 AC109272	AC109272 Mus muscu
C 9	167.6	25.7	251206	2 AC107815	AC107815 Mus muscu
C 10	95.4	14.7	136098	9 AC006970	AC006970 Homo sapi
C 11	93.8	14.4	51891	2 AC074103	AC074103 Homo sapi
C 12	93.4	14.3	678	9 BC009696	BC009696 Homo sapi
C 13	93.4	14.3	905	6 AX337050	AX337050 Sequence
C 14	93.4	14.3	905	9 HS18D	HS18D Human 1-8D
C 15	93.2	14.3	402	9 BT006892	BT006892 Homo sapi
C 16	93.2	14.3	402	12 BT007876	BT007876 Synthetic
C 17	93.2	14.3	637	9 BC006794	BC006794 Homo sapi
C 18	93.2	14.3	642	6 AR225472	AR225472 Sequence
C 19	93.2	14.3	642	6 AX321588	AX321588 Sequence
C 20	93.2	14.3	645	9 BC022439	BC022439 Homo sapi
C 21	93.2	14.3	695	6 AX281855	AX281855 Sequence
C 22	93.2	14.3	749	9 BC008417	BC008417 Homo sapi
C 23	93.2	14.3	808	9 HS18U	HS18U Human 1-8U
C 24	91.6	14.1	37918	9 AP005232	AP005232 Homo sapi
C 25	91.6	14.1	222765	2 AC144988	AC144988 Gorilla g
C 26	91.2	14.0	378	9 BT007173	BT007173 Homo sapi
C 27	91.2	14.0	378	12 BT007946	BT007946 Synthetic
C 28	91.2	14.0	647	9 HSNR827	HSNR827 H. sapiens m
C 29	91.2	14.0	683	9 BC000897	BC000897 Homo sapi
C 30	91.2	14.0	842	6 AX397528	AX397528 Sequence
C 31	91.2	14.0	851	6 AX014337	AX014337 Sequence
C 32	91.2	14.0	853	6 AX302539	AX302539 Sequence
C 33	91.2	14.0	853	6 AX409461	AX409461 Sequence
C 34	91.2	14.0	853	6 HM927A	HM927A Human inter
C 35	91.2	14.0	1095	6 AX281854	AX281854 Sequence
C 36	91.2	14.0	245	6 AX340862	AX340862 Sequence
C 37	90.8	13.9	579	4 AR272041	AR272041 Bos tauru
C 38	90.8	13.9	107717	9 AC137055	AC137055 Homo sapi
C 39	90.8	13.9	110000	2 AC004085_1	AC004085_1 Continuation (2 of
C 40	90.8	13.9	110000	2 AC004085_2	AC004085_2 Continuation (3 of
C 41	90.2	13.9	123351	9 AC023157	AC023157 Homo sapi
C 42	89.6	13.8	905	10 BC027285	BC027285 Mus muscu
C 43	89.6	13.8	84979	9 AC116049	AC116049 Homo sapi
C 44	89.6	13.8	176104	2 AC016193	AC016193 Homo sapi
C 45	89.6	13.8	189572	2 AC084812	AC084812 Homo sapi

#### ALIGNMENTS

RESULT 1  
AF015416/c 95038 bp DNA linear PRI 20-AUG-1997  
LOCUS Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.  
DEFINITION AF015416  
ACCESSION AF015416.1 GI:2335202  
VERSION  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
REFERENCE  
AUTHORS  
Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davis,J.,  
Davies,C.U., Davis,C., English,C., Fondon,T., Franklin,T.L.,  
Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,

Pred. No. is the number of results predicted by chance to have a

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K.,  
Olliver, T., Patel, P., Prober, S., Rayner, S., Schageman, J.,  
Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T.,  
Wilson, R. and Burbee, D.  
95 Kb from Four Overlapping Human Chromosome 11p15.5 Cosmids  
Unpublished  
2 (bases 1 to 95038)  
Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Davie, J.,  
Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L.,  
Gamer, H.R., Gordon, M., Grant, O., Hahner, L., Harris, J.,  
Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K.,  
Olliver, T., Patel, P., Prober, S., Rayner, S., Schageman, J.,  
Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T.,  
Wilson, R. and Burbee, D.  
Direct Submission  
Submitted (18-JUL-1997) Genome Science and Technology Center,  
University of Texas Southwestern Medical Center at Dallas, 5323  
Harry Hines Blvd, Dallas, TX 75235-8591, USA

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K.,  
Olliver, T., Patel, P., Prober, S., Rayner, S., Schageman, J.,  
Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T.,  
Wilson, R. and Burbee, D.  
95 Kb from Four Overlapping Human Chromosome 11p15.5 Cosmids  
Unpublished  
2 (bases 1 to 95038)  
Evans, G.A., Bradbury, P., Brignac, S., Bumeister, R., Davie, J.,  
Davies, C.J., Davis, C., English, C., Fondon, T., Franklin, T.L.,  
Gamer, H.R., Gordon, M., Grant, O., Hahner, L., Harris, J.,  
Hinson, S., Megarity, C., Narayanaswamy, U., Newton, J., O'Brien, K.,  
Olliver, T., Patel, P., Prober, S., Rayner, S., Schageman, J.,  
Schilling, P., Schultz, R., Syed, M., Valenzuela, D., Ward, T.,  
Wilson, R. and Burbee, D.  
Direct Submission  
Submitted (18-JUL-1997) Genome Science and Technology Center,  
University of Texas Southwestern Medical Center at Dallas, 5323  
Harry Hines Blvd, Dallas, TX 75235-8591, USA

FEATURES  
SOURCE  
1. .95038  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11p15.5"  
711. .975  
repeat\_region /rpt\_family="Alu"  
repeat\_region 2728. .3022  
/rpt\_family="Alu"  
repeat\_region 3363. .3719  
/rpt\_family="Alu"  
repeat\_region 5634. .5909  
/rpt\_family="Alu"  
repeat\_region complement(6720. .6959)  
/rpt\_family="Alu"  
repeat\_region 10560. .10839  
/rpt\_family="Alu"  
repeat\_region complement(11832. .12455)  
/rpt\_family="Alu"  
repeat\_region 12921. .13405  
/rpt\_family="Alu"  
repeat\_region 13718. .14314  
/rpt\_family="Alu"  
repeat\_region 14617. .14779  
/rpt\_family="Alu"  
repeat\_region 15396. .15695  
/rpt\_family="Alu"  
repeat\_region 15722. .15994  
/rpt\_family="Alu"  
repeat\_region complement(16696. .16997)  
/rpt\_family="Alu"  
repeat\_region complement(18996. .19286)  
/rpt\_family="Alu"  
repeat\_region complement(19458. .19747)  
/rpt\_family="Alu"  
repeat\_region complement(19777. .20045)  
/rpt\_family="Alu"  
repeat\_region complement(22885. .23002)  
/rpt\_family="Alu"  
repeat\_region 23109. .23361  
/rpt\_family="Alu"  
repeat\_region 23292. .23481  
/rpt\_family="Alu"  
repeat\_region complement(23877. .23944)  
/rpt\_family="Alu"  
repeat\_region complement(24695. .24986)  
/rpt\_family="Alu"  
repeat\_region complement(25429. .26012)  
/rpt\_family="Alu"  
repeat\_region complement(27380. .27676)  
/rpt\_family="Alu"  
repeat\_region complement(29422. .29486)  
/rpt\_family="Alu"

repeat\_region complement(29514. .29802)  
/rpt\_family="Alu"  
repeat\_region 30797. .31075  
/rpt\_family="Alu"  
repeat\_region 31688. .31792  
/rpt\_family="Alu"  
repeat\_region 32125. .32412  
/rpt\_family="Alu"  
repeat\_region complement(36871. .39186)  
/rpt\_family="Alu"  
repeat\_region 42952. .43094  
/rpt\_family="Alu"  
repeat\_region 43106. .43388  
/rpt\_family="Alu"  
repeat\_region 43754. .44216  
/rpt\_family="Alu"  
repeat\_region complement(44815. .45136)  
/rpt\_family="Alu"  
repeat\_region 45145. .45484  
/rpt\_family="Alu"  
repeat\_region 45740. .45995  
/rpt\_family="Alu"  
repeat\_region 46375. .46646  
/rpt\_family="Alu"  
repeat\_region 47145. .47738  
/rpt\_family="Alu"  
repeat\_region 48353. .48626  
/rpt\_family="Alu"  
repeat\_region 49567. .49842  
/rpt\_family="Alu"  
repeat\_region 49940. .50211  
/rpt\_family="Alu"  
repeat\_region complement(50615. .50907)  
/rpt\_family="Alu"  
repeat\_region complement(53001. .53290)  
/rpt\_family="Alu"  
repeat\_region complement(54161. .54512)  
/rpt\_family="Alu"  
repeat\_region 54533. .54823  
/rpt\_family="Alu"  
repeat\_region 57899. .58113  
/rpt\_family="Alu"  
repeat\_region complement(58795. .59044)  
/rpt\_family="Alu"  
repeat\_region complement(59414. .60023)  
/rpt\_family="Alu"  
repeat\_region complement(60177. .60311)  
/rpt\_family="Alu"  
repeat\_region complement(60770. .61108)  
/rpt\_family="Alu"  
repeat\_region complement(61207. .61557)  
/rpt\_family="Alu"  
repeat\_region 61670. .61942  
/rpt\_family="Alu"  
repeat\_region 61944. .62436  
/rpt\_family="Alu"  
repeat\_region complement(67807. .68105)  
/rpt\_family="Alu"  
repeat\_region 70961. .71049  
/rpt\_family="Alu"  
repeat\_region 71224. .71523  
/rpt\_family="Alu"  
repeat\_region complement(85033. .85397)  
/rpt\_family="Alu"  
repeat\_region complement(85414. .85620)  
/rpt\_family="Alu"  
repeat\_region complement(85701. .85804)  
/rpt\_family="Alu"  
repeat\_region complement(85895. .86028)  
/rpt\_family="Alu"  
repeat\_region 86180. .86477  
/rpt\_family="Alu"  
repeat\_region 86497. .86780





Db 82847 TAACCGAGATTCCCGGCGCTACTGTGCC 82815

RESULT 3  
AP006283  
LOCUS  
DEFINITION  
AP006283 156819 bp DNA linear HTG 26-MAR-2003  
Homo sapiens chromosome 11 clone RP11-326C3 map 11p, WORKING DRAFT  
SEQUENCE, 19 unordered pieces.

ACCESSION  
AP006283  
VERSION  
AP006283.1 GI:29243343  
KEYWORDS  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE  
AUTHORS  
1 Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE  
JOURNAL  
Homo sapiens genomic DNA of 11p  
REFERENCES  
2 Published Only in Database (2003)  
2 (bases 1 to 156819)  
Hattori, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,  
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.  
TITLE  
JOURNAL  
Direct Submission  
Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical  
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),  
1-7-22, Suenitro-chou, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
(E-mail: hattori@psc.riken.go.jp, URL: http://hsp.gsc.riken.go.jp/  
Tel: 81-45-503-9111, Fax: 81-45-503-9170)

COMMENT  
----- Genome Center  
Center: RIKEN Genomic Sciences Center (GSC)  
Center code: RIKEN  
Web site: http://hsp.gsc.riken.go.jp/  
Contact: hattori@psc.riken.go.jp  
----- Project Information  
Center project name: HumDraft11  
Center clone name: RP11-326C3  
----- Summary Statistics  
Sequencing vector: PCR products; 100% of reads  
Chemistry: Dye-terminator ET-amersham; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 144706 bases at least Q40  
Consensus quality: 150967 bases at least Q30  
Consensus quality: 153657 bases at least Q20  
Insert size: 155019; sum-of-contigs  
Quality coverage: 6.36x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of  
19 contigs. The true order of the pieces is not known and their  
order in this sequence record is arbitrary. Gaps between the  
contigs are represented as runs N, but the exact sizes of the gaps  
are unknown. This record will be updated with the finished sequence  
as soon as it is available and the accession number will be  
preserved

1 21001 contig of 21001 bp in length  
21102 40714 contig of 19613 bp in length  
40815 58430 contig of 17616 bp in length  
58531 58430 contig of 8224 bp in length  
66855 81134 contig of 14280 bp in length  
81235 89763 contig of 8529 bp in length  
89864 101889 contig of 12026 bp in length  
101990 112726 contig of 10636 bp in length  
112726 119802 contig of 7077 bp in length  
119802 125181 contig of 5279 bp in length  
125181 125282 contig of 3903 bp in length  
125282 137260 contig of 7976 bp in length  
137260 143542 contig of 6182 bp in length  
143542 146891 contig of 3249 bp in length  
146891 149568 contig of 2577 bp in length  
149568 149668 contig of 100 bp in length  
149668 152678 contig of 3010 bp in length  
152678 154135 contig of 1357 bp in length  
154135 154235 contig of 100 bp in length  
154235 155662 contig of 1427 bp in length  
155662 155762 contig of 100 bp in length  
155762 156819 contig of 1057 bp in length

NOTE: This is a 'working draft' sequence. It currently  
\* consists of 19 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 21001: contig of 21001 bp in length  
21002 21101: gap of 100 bp  
21102 40714: contig of 19613 bp in length  
40715 40814: gap of 100 bp  
40815 58430: contig of 17616 bp in length  
58431 58530: gap of 100 bp  
58531 66854: contig of 8224 bp in length  
66855 66854: gap of 100 bp  
66854 81134: contig of 14280 bp in length  
81135 81234: gap of 100 bp  
81235 89763: contig of 8529 bp in length  
89764 89863: gap of 100 bp  
89864 101889: contig of 12026 bp in length  
101890 101889: gap of 100 bp  
112625: contig of 10636 bp in length  
112725: gap of 100 bp  
112726 119802: contig of 7077 bp in length  
119803 119902: gap of 100 bp  
119903 125181: contig of 5279 bp in length  
125182 125281: gap of 100 bp  
125282 129184: contig of 3903 bp in length  
129185 129284: gap of 100 bp  
129285 137260: contig of 7976 bp in length  
137261 137360: gap of 100 bp  
137361 143542: contig of 6182 bp in length  
143543 143642: gap of 100 bp  
143643 146891: contig of 3249 bp in length  
146892 146991: gap of 100 bp  
146992 149568: contig of 2577 bp in length  
149569 149668: gap of 100 bp  
149669 152678: contig of 3010 bp in length  
152679 152778: gap of 100 bp  
152779 154135: contig of 1357 bp in length  
154136 154235: gap of 100 bp  
154236 155662: contig of 1427 bp in length  
155663 155762: gap of 100 bp  
155763 156819: contig of 1057 bp in length.

FEATURES  
source  
1..156819  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11p"  
/clone="RP11-326C3"  
1..21001  
/note="assembly\_fragment"  
21102..40714  
/note="assembly\_fragment"  
40815..58430  
/note="assembly\_fragment"  
58531..66754  
/note="assembly\_fragment clone\_end:Sp6 vector\_side:left"  
66855..81134  
/note="assembly\_fragment"  
81235..89763  
/note="assembly\_fragment"  
89864..101889  
/note="assembly\_fragment"  
101890..112625  
/note="assembly\_fragment"  
112726..119802  
/note="assembly\_fragment"  
119903..125181  
/note="assembly\_fragment"

```

misc_feature      125282..129184
                  /note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature      129285..137260
                  /note="assembly_fragment"
misc_feature      137361..143542
                  /note="assembly_fragment"
misc_feature      143643..146891
                  /note="assembly_fragment"
misc_feature      146992..149368
                  /note="assembly_fragment"
misc_feature      149669..152678
                  /note="assembly_fragment"
misc_feature      152779..154135
                  /note="assembly_fragment"
misc_feature      154236..155662
                  /note="assembly_fragment"
misc_feature      155763..156819
                  /note="assembly_fragment"
BASE COUNT      37089 a 39987 c 40186 g 37757 t 1800 others
ORIGIN
Query Match      68.6%; Score 446.6; DB 2; Length 156819;
                  Best Local Similarity 99.1%; Pred.No.2.3e-78;
Matches 449; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY      199 CTCATCAAGCCCGGAGTCAAGAGGTGTGTGACCTGGAAGCGGCCGCTTTGG 258
DB      92819 CTCCTGAGGCGCGAGATCGAAGGTGTGTGACCTGGAAGCGGCCGCTTTGG 92878
QY      259 CTCGAAGCCGAAGTCTCAACATCTGCGCGAGTGTGAGCGTGGTCCGCACTGCT 318
DB      92879 CTCGAAGCCGAAGTCTCAACATCTGCGCGAGTGTGAGCGTGGTCCGCACTGCT 92938
QY      319 GCTCTGGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 378
DB      92939 GCTCTGGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 92998
QY      379 GCGCTTCTTCAGCAACAGTTTGTATGAGCGGACTATGACAGGTGGTGTCTGATC 438
DB      92999 GCGCTTCTTCAGCAACAGTTTGTATGAGCGGACTATGACAGGTGGTGTCTGATC 93058
QY      439 TGGGGGCACTAGCCCGGAGCACTGACCCAGGCTGTGCGCCCTGGGGCCCAATACGACT 498
DB      93059 TGGGGGCACTAGCCCGGAGCACTGACCCAGGCTGTGCGCCCTGGGGCCCAATACGACT 93118
QY      499 CCGCGGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 558
DB      93119 CCGCGGAGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 93178
QY      559 CCGTTCAGCCCGGAGCACTGAGGCTGAAACAGTCAAGCCCGGGGTCTTCAACC 618
DB      93179 CCGTTCAGCCCGGAGCACTGAGGCTGAAACAGTCAAGCCCGGGGTCTTCAACC 93238
QY      619 TAACCGGAGGTCCCGGGCCCTAATCTGCGC 651
DB      93239 TAACCGGAGGTCCCGGGCCCTAATCTGCGC 93271

RESULT 4
AP006286/c      175416 bp DNA linear HTG 26-MAR-2003
LOCUS
DEFINITION      Homo sapiens chromosome 11 clone RP13-25N22 map 11p, WORKING DRAFT
SEQUENCE
ACCESSION      AP006286.1 GI:29243346
VERSION
KEYWORDS      HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Toroki,Y., Watanabe,H. and Sakaki,Y.

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Homo sapiens genomic DNA of 11p
Published Only in Database (2003)
2 (bases 1 to 175416)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Toroki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-45-503-9111, Fax:81-45-503-9170)
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gsc.riken.go.jp
----- Project Information
Center project name: Hmndraft11
Center clone name: RP13-25N22
----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 160074 bases at least Q40
Consensus quality: 167722 bases at least Q30
Consensus quality: 171662 bases at least Q20
Insert size: 173316; sum-of-ctrls
Quality coverage: 7.92x in Q20 bases; sum-of-ctrls
-----
NOTE: This is a 'working draft' sequence. It currently consists of
22 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved
1 33331 contig of 33331 bp in length
33232 54098 contig of 17869 bp in length
54199 72037 contig of 20867 bp in length
72138 83277 contig of 11140 bp in length
83378 98599 contig of 15222 bp in length
98700 105513 contig of 16614 bp in length
105614 116229 contig of 10616 bp in length
116330 121964 contig of 5635 bp in length
122065 127627 contig of 4544 bp in length
127728 131971 contig of 4244 bp in length
132072 137506 contig of 5435 bp in length
137607 143829 contig of 6223 bp in length
143930 148539 contig of 4610 bp in length
148640 153148 contig of 4509 bp in length
153249 157509 contig of 4261 bp in length
157610 160561 contig of 2952 bp in length
160662 164778 contig of 4117 bp in length
164879 168269 contig of 3391 bp in length
168370 170689 contig of 2220 bp in length
170790 172833 contig of 2044 bp in length
172934 174155 contig of 1222 bp in length
174256 175416 contig of 1161 bp in length.
NOTE: This is a 'working draft' sequence. It currently
consists of 22 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 33331: contig of 33331 bp in length
33231: gap of 100 bp
33232 54098: contig of 20867 bp in length
54099 72037: gap of 100 bp
72038 83277: contig of 17839 bp in length
83278 98599: gap of 100 bp
98600 105513: contig of 11140 bp in length

```

```

*      83378      83377: gap of 100 bp
*      83378      98599: contig of 15222 bp in length
*      98600      98600: gap of 100 bp
*      98700      105513: contig of 6814 bp in length
*      105514      105613: gap of 100 bp
*      105614      116329: contig of 10616 bp in length
*      116330      121964: gap of 100 bp
*      121965      122064: gap of 100 bp
*      122065      127627: contig of 5563 bp in length
*      127628      131971: contig of 4244 bp in length
*      131972      132071: gap of 100 bp
*      132072      137506: contig of 5435 bp in length
*      137507      137607: gap of 100 bp
*      137607      143829: contig of 6223 bp in length
*      143830      143929: gap of 100 bp
*      143930      148539: contig of 4610 bp in length
*      148540      153148: gap of 100 bp
*      148640      153248: gap of 100 bp
*      153249      157509: contig of 4261 bp in length
*      157510      157609: gap of 100 bp
*      157610      160651: contig of 2952 bp in length
*      160652      160661: gap of 100 bp
*      160662      164778: contig of 4117 bp in length
*      164779      164879: gap of 100 bp
*      164879      168269: contig of 3391 bp in length
*      168270      168369: gap of 100 bp
*      168370      170689: contig of 2320 bp in length
*      170690      170789: gap of 100 bp
*      170790      172833: contig of 2044 bp in length
*      172834      172933: gap of 100 bp
*      172934      174155: contig of 1222 bp in length
*      174156      174255: gap of 100 bp
*      174256      175416: contig of 1161 bp in length.

```

## FEATURES

```

source
1. .175416
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="11"
/map="11p"
/clone="RP13-25N22"
1. .33131
/note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature
33232..54098
/note="assembly_fragment"
misc_feature
54199..72037
/note="assembly_fragment"
misc_feature
72138..83277
/note="assembly_fragment"
misc_feature
83378..98599
/note="assembly_fragment"
misc_feature
98700..105513
/note="assembly_fragment clone_end:T7 vector_side:left"
misc_feature
105614..116229
/note="assembly_fragment"
misc_feature
116330..121964
/note="assembly_fragment"
misc_feature
122065..127627
/note="assembly_fragment"
misc_feature
127728..131971
/note="assembly_fragment"
misc_feature
132072..137506
/note="assembly_fragment"
misc_feature
137607..143829
/note="assembly_fragment"
misc_feature
143930..148539
/note="assembly_fragment"
misc_feature
148640..153148
/note="assembly_fragment"
misc_feature
153249..157509
/note="assembly_fragment"

```

```

/note="assembly_fragment"
157610..160561
/note="assembly_fragment"
160662..164778
/note="assembly_fragment"
164879..168269
/note="assembly_fragment"
168370..170689
/note="assembly_fragment"
170790..172833
/note="assembly_fragment"
172934..174155
/note="assembly_fragment"
174256..175416
/note="assembly_fragment"
BASE COUNT 37167 a 49392 c 49417 g 37339 t 2101 others
ORIGIN

```

```

Query Match 68.6%; Score 446.6; DB 2; Length 175416;
Best Local Similarity 99.1%; Pred. No. 2.2e-78;
Matches 449; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

QY 199 CTCATCAAGGCCCGAATCAAGTGTGTGACCTCGAAGCGGCCGGCTTTGG 258
DB 42209 CTCCTGCAAGGCCCGAATCAAGTGTGTGACCTCGAAGCGGCCGGCTTTGG 42150
QY 259 CTCCAAGCCCAAGTGTCTCAATCATCTGCGCCGATGTGAGCGTGTGCGCCCACTGCT 318
DB 42149 CTCCAAGCCCAAGTGTCTCAATCATCTGCGCCGATGTGAGCGTGTGCGCCCACTGCT 42090
QY 319 GCTCTGCGGGCTGTGTGTGACTGTGTGCTGCACTGCGCCGCTGCGCCCAAGACTTGC 378
DB 42089 GCTCTGCGGGCTGTGTGTGACTGTGTGCTGCACTGCGCCGCTGCGCCCAAGACTTGC 42030
QY 379 CGCCTTCTTCAACCAAGTGTGATGACCGCGGACATATGACATACGCGTGTGCTGATC 438
DB 42029 CGCCTTCTTCAACCAAGTGTGATGACCGCGGACATATGACATACGCGTGTGCTGATC 41970
QY 439 TGGGGGACATAGCCCGAAGACATGACCCGAGGCTGCTGCGCCGCTGCGCCCAATACACT 498
DB 41969 TGGGGGACATAGCCCGAAGACATGACCCGAGGCTGCTGCGCCGCTGCGCCCAATACACT 41910
QY 499 CCCCAGAGCTGAGCCCTCTCTGTGTGGGGCTTCATCCCTGCCCCATCCTGATCTGAGG 558
DB 41909 CCCCAGAGCTGAGCCCTCTCTGTGTGGGGCTTCATCCCTGCCCCATCCTGATCTGAGG 41850
QY 559 CCCTGACGCCCAATATGAGGACCTTAAGCTTAACACAGACACACCCGGGGCTTCAACC 618
DB 41849 CCCTGACGCCCAATATGAGGACCTTAAGCTTAACACAGACACACCCGGGGCTTCAACC 41790
QY 619 TAACCGAGAGTCCCGGGCCCTTAACCTGAGCC 651
DB 41789 TAACCGAGAGTCCCGGGCCCTTAACCTGAGCC 41757

```

```

RESULT 5
AC138230/c 187160 bp DNA linear PRI 22-MAY-2003
LOCUS Homo sapiens chromosome 11, clone RP13-317D12, complete sequence.
DEFINITION AC138230
ACCESSION AC138230.5 GI:30984765
VERSION AC138230.5 GI:30984765
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187160)
AUTHORS Birren,B., Nuebaum,C. and Lander,E.
TITLE Homo sapiens chromosome 11, clone RP13-317D12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187160)
AUTHORS Birren,B., Nuebaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Baerlien,V., Bloom,T., Bogunlavskiy,L., Boukigalter,B.,

```



Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlynga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (20-DEC-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 187160)

Bitren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, T., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlynga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

4 (bases 1 to 187160)

Bitren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Boguslavsky, L., Bouhgalter, B., Camarata, J., Chang, J., Choepel, Y., Collamore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K., Diaz, J.S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafez, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C., Macdonald, P., Major, J., Manning, T., Matthews, C., McCarthy, M., Meldrim, J., Menus, L., Mihova, T., Mlynga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhag, P., Pierre, N., Rachupka, A., Ramsamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Venkataraman, V.S., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (22-MAY-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On May 22, 2003 this sequence version replaced g1:29423936.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L29013  
Center clone name: 317\_D\_12

----- Location/Qualifiers

1. 187160  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="11"  
/map="11"  
/clone="RP13-317D12"  
/clone\_1kb="RP13-317D12 Human Female BAC"

repeat\_region  
135..700  
/rpt\_family="L2"  
repeat\_region  
1134..1310  
/rpt\_family="MIR"  
complement(1311..1614)  
/rpt\_family="ALUy"  
repeat\_region  
1615..1638  
/rpt\_family="MIR"  
repeat\_region  
1849..1938  
/rpt\_family="L1ME"  
complement(4297..4385)  
/rpt\_family="L2"  
repeat\_region  
4463..4561  
/rpt\_family="FRAM"  
complement(4573..4613)  
/rpt\_family="MADE1"  
repeat\_region  
4733..5040  
/rpt\_family="ALUy"  
complement(5094..5383)  
/rpt\_family="L2"  
repeat\_region  
5379..5505  
/rpt\_family="L2"  
complement(5501..5627)  
/rpt\_family="L2"  
repeat\_region  
5642..6096  
/rpt\_family="L2"  
repeat\_region  
6342..6666  
/rpt\_family="MER41E"  
repeat\_region  
6639..6719  
/rpt\_family="GC-rich"  
repeat\_region  
6917..6996  
/rpt\_family="C-rich"  
repeat\_region  
7136..7329  
/rpt\_family="C-rich"  
repeat\_region  
15461..15483  
/rpt\_family="(TG)n"  
repeat\_region  
18178..18413  
/rpt\_family="L2"  
complement(18543..18907)  
/rpt\_family="MER4A"  
complement(18913..19138)  
/rpt\_family="ALUSg/x"  
complement(19139..19318)  
/rpt\_family="MER4D"  
complement(19319..19688)  
/rpt\_family="MER4E1"  
19689..19992  
/rpt\_family="ALUSp"  
repeat\_region  
19994..20308  
/rpt\_family="ALUSg"  
complement(20309..20630)  
/rpt\_family="MER4E1"  
repeat\_region  
20632..20763  
/rpt\_family="ALUSg/x"  
complement(20785..20882)  
/rpt\_family="MER4E"  
complement(20886..21091)

```

repeat_region /rpt_family="MER41E"
                21092..21402
repeat_region /rpt_family="AluSx"
                complement(21403..21580)
repeat_region /rpt_family="MER41E"
                complement(22261..22538)
repeat_region /rpt_family="MER41E"
                complement(22543..22594)
repeat_region /rpt_family="MER4D"
                22595..22902
repeat_region /rpt_family="AluY"
                complement(22903..23195)
repeat_region /rpt_family="MER4D"
                23197..23314
repeat_region /rpt_family="FLAM_C"
                complement(23318..23368)
repeat_region /rpt_family="MER4D"
                23369..23890
repeat_region /rpt_family="MER4A"
                complement(23891..24166)
repeat_region /rpt_family="MER4D"
                complement(24381..24698)
repeat_region /rpt_family="MER63"
                complement(25372..25668)
repeat_region /rpt_family="AluY"
                26402..26493
repeat_region /rpt_family="(TG)n"
                complement(27325..28044)
repeat_region /rpt_family="LTR30"
                28747..28755
unsure /note="<30 qual SNGI region"
repeat_region 29024..29047
repeat_region /rpt_family="GC_rich"
                complement(29397..30103)
repeat_region /rpt_family="LTR8"
                30537..30656
repeat_region /rpt_family="LTR2C"
                complement(30657..31083)
repeat_region

Query Match 68.6%; Score 446.6; DB 9; Length 187160;
Best Local Similarity 99.1%; Pred. No. 2.2e-78;
Matches 449; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 199 CTCACATCAGGCGCCGAGATCAGAGGTGTGTGTGACTGTGAACGCGCCGCGGCTTTTGG 258
Db 16606 CTCCTGTGAGGCGCCGAGATCAGAGGTGTGTGTGACTGTGAAGCGCGCCGCGGCTTTTGG 16547
QY 259 CTCCAAAGCCAAAGTGTACACATCTGTGCGCGGATGTGAGCGTGTGCGCCGACATGCT 318
Db 16546 CTCCAAAGCCAAAGTGTACACATCTGTGCGCGGATGTGAGCGTGTGCGCCGACATGCT 16487
QY 319 GCTCTGTGGGCTGTGTGTGTGACTGTGCTGTGCACTGTGCGCGGCTGTGCGCAAGACCTTGC 378
Db 16486 GCTCTGTGGGCTGTGTGTGTGACTGTGCTGTGCACTGTGCGCGGCTGTGCGCAAGACCTTGC 16427
QY 379 GGCCTTCTCAGACCAAAATTGTGATGACCGGACCTTGTGACTGACAGCTGTGCTGTGATC 438
Db 16426 GGCCTTCTCAGACCAAAATTGTGATGACCGGACCTTGTGACTGACAGCTGTGCTGTGATC 16367
QY 439 TGGGGGCACTAGGCGCCAGACACTGACCCGAGGCTGTGCGCCGCGGCGCCCAATAGTACT 498
Db 16366 TGGGGGCACTAGGCGCCAGACACTGACCCGAGGCTGTGCGCCGCGGCGCCCAATAGTACT 16307
QY 499 CCCCAGAGCTGTGCGCTCTTCTGTGTGGGCTTGCATCTCTGCTGCTGATCTGTGAGG 558
Db 16306 CCCCAGAGCTGTGCGCTCTTCTGTGTGGGCTTGCATCTCTGCTGCTGATCTGTGAGG 16247
QY 559 CCCTCAAGCCCAACAATGGGACCTTAAGGCTGAACAAGTCAAGACCCCGGGGCTTTTACC 618
Db 16246 CCCTCAAGCCCAACAATGGGACCTTAAGGCTGAACAAGTCAAGACCCCGGGGCTTTTACC 16187
QY 619 TAACCCGAGAGTTCGCGGCGCCCTTAATCTGTGCC 651

```

```

Db 16186 TAACCCGAGAGTTCGCGGCGCCCTTAATCTGTGCC 16154

RESULT 6
LOCUS MMU9781 499 bp mRNA linear ROD 22-MAR-2001
DEFINITION Mus musculus mRNA for putative haemopoietic membrane protein.
ACCESSION AJ009781.1 GI:3355654
VERSION haemopoietic membrane protein; putative.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE
AUTHORS Baird,J.W., Ryan,K.M., Hayes,I., Hampson,L., Heyworth,C.M.,
Clark,A., Woolton,M., Ansell,J.D., Menzel,U., Hole,N. and
Graham,G.J.
TITLE Differentiating embryonal stem cells are a rich source of
JOURNAL haemopoietic gene products and suggest erythroid preconditioning of
MEDLINE primitive haemopoietic stem cells
PUBMED J. Biol. Chem. 276 (12), 9189-9198 (2001)
11106657
2 (bases 1 to 499)
REFERENCE
AUTHORS Graham,G.J.
TITLE Direct Submision
JOURNAL Submitted (24-JUL-1998) Graham G.J., CRC Beatson Laboratories,
Beatson Institute for Cancer Research, Gartcube Estate, Switchback
Road, Bearsden, Glasgow G61 1BD, SCOTLAND, UK
FEATURES
source
1..499
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="EPG1"
/cell_line="ES"
/dev_stage="d5 of in vitro development"
31..436
/note="ORF1"
/codon_start=1
/product="putative haemopoietic membrane protein"
/protein_id="CA08829.1"
/db_xref="GI:3355655"
/db_xref="SPTREMBL:O88728"
/translation="MDTSYPRDPPRAPSRRKDAAHATLSMTGPTPRDMLWSVF
STMYNLCCIGFLAVHSVKARDQKAGLBARQYGSVAKXCNILAAWMTLVPELL
LGLVVTGAHLHSKIAKDAAPSTFDEBDYV"
BASE COUNT 107 a 155 c 137 g 100 t
ORIGIN

Query Match 43.7%; Score 284.6; DB 10; Length 499;
Best Local Similarity 79.6%; Pred. No. 4.8e-46;
Matches 351; Conservative 0; Mismatches 84; Indels 6; Gaps 1;

QY 1 GGAAGAGACGCGCTGTGAAACCATGACAGCGCGTATCCCGCGAGACACCCGCGCC 60
Db 9 GGAAGAGACGCGCTGTGAAACCATGACAGCGCGTATCCCGCGAGACACCCGCGCTCC 68
QY 61 CAGCGCCGACGAGGCGGT-----GCCACACAGCCCTCAACATGAGCGCGCCCGACCC 114
Db 69 ATCATCCGCAAGGCTGATGTGACAGCCACAGCCCTCTCATAGGAAACCTGTGGCC 128
QY 115 CCGGCTCGAGACCACTTGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 174
Db 129 TACACAGAGATTCATGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 188
QY 175 CCTGGCTTCTGTGGGCTGTGCTTCAATCAAGGCGGAGATCAAGAGTGTGTGTGTGTGT 234
Db 189 CCTGTGATTCCTGTGGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 248
QY 235 CCTGGAAGGCGCGGCGGCTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 294

```

Db		249	CTTGAGGCGTCAAGAGGCAGTAGTMGGTCACAAAGCCAAAGTGCTAACACATCCTGGCGTGCAT	308
Oy		295	GTGACGCGTGGGCCGACCATGCTCTCTCTCTGGGAGCTGGTGTACTGGTCCCTGCACCT	354
Db		309	GTTGACATATTGGTGCCCCCATGTGCTGCTCTCTGGACCTGGTGGAATGCGCGCTTGACCT	368
Oy		355	GGCCCGGCTGGCCAAAGACCTGCGCGCTTCTTTCAGCACCAAGTTTATNTAGAAGGAGACTA	414
Db		369	GTTCAAAGTTAAGCCAAAGACTCTGGCGGCTTTTTCAGCACCAAGTTTATNTAGAAGAGACTA	428
Oy		415	TGACTGACAGCGCTGGGCTCTG	435
Db		429	TAACTAAGATTTCCGAGCCTG	449
RESULT 7				
AC132968/c				
LOCUS				
DEFINITION			Rattus norvegicus clone CH230-172C3, *** SEQUENCING IN PROGRESS	
ACCESSION			AC132968	
VERSION			AC132968.4	
KEYWORDS			HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.	
SOURCE			Rattus norvegicus (Norway rat)	
ORGANISM			Rattus norvegicus	
REFERENCE			Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
AUTHORS			Allen, D., Marie, Metzker, M., Lee, Abramzon, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Ayagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandartanaik, D., Barber, M., Barnstead, M., Bernamed, F., Biewald, K., Blair, J., Blankenburg, K., Bluth, P., Brown, M., Bryant, N., Buhay, C., Burck, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., L., Davis, C., Davy-Carroll, L., De Andrade, C., Dederich, D., Delgado, O., Denison, S., Deremo, C., Ding, Y., Dinh, H., Divya, K., Diaper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duval, B., Evans, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falle, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Georgerogios, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guayra, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howells, S., Huliyk, S., Hunne, V., Idaldi, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Kapachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C., Kowis, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenzuewa, L., Louisseg, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mamoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mathewey, S., McLeod, M.P., McNell, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minelli, A.E., Montemeyer, J., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Natr, L., Narkervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaoketemen, O., Okwonu, G., Olarnpunaagoon, A., Pal, S., Parke, K., Patelant, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plapper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., -L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Kelly, B., Kelly, W., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodney, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sismon, I., Slitter, C.D., Smys, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steidle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tinney, A., Trojes, Z., Umani, K., Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,	

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Unpublished
2 (bases 1 to 238533)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (05-SEP-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 238533)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:22264838.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KBMS
Center clone name: CH230-172C3
----- Summary Statistics
Assembly program: Phrap, version 0.990329
Consensus quality: 225317 bases at least Q40
Consensus quality: 225463 bases at least Q30
Consensus quality: 226655 bases at least Q20
Estimated insert size: 230654; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbankdraft\_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
222461: contig of 222461 bp in length
222462
222561: gap of unknown length
222562
222949: contig of 2388 bp in length
224950
225049: gap of unknown length
225050
228242: contig of 3193 bp in length
228243
228342: gap of unknown length
228343
238533: contig of 10191 bp in length.

Location/Qualifiers
1. 238533
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-172C3"
1. 2349
/note="wgs_contig"

misc_feature

```



FEATURES	SOURCE
*	109183 109282: gap of 100 bp
*	109283 133485: contig of 24703 bp in length
*	133486 133585: gap of 100 bp
*	133586 157263: contig of 23578 bp in length
*	157264 157363: gap of 100 bp
*	157364 192898: contig of 35535 bp in length
*	192899 200734: gap of 100 bp
*	192999 200734: contig of 7736 bp in length.
location/Qualifiers	
1..200734	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/db_xref="taxon:10090"
	/clone="RP23-354I18"
	/clone_lib="RPCI-23 Female Mouse BAC"
1..5430	/note="assembly_fragment"
	clone_end:SP6
	vector_side:left"
5531..6612	/note="assembly_fragment"
6713..8249	/note="assembly_fragment"
8350..10942	/note="assembly_fragment"
11043..14087	/note="assembly_fragment"
14188..20506	/note="assembly_fragment"
20607..26335	/note="assembly_fragment"
26436..31423	/note="assembly_fragment"
31524..39064	/note="assembly_fragment"
39165..47890	/note="assembly_fragment"
47891..56258	/note="assembly_fragment"
56359..66903	/note="assembly_fragment"
67004..109182	/note="assembly_fragment"
109283..133485	/note="assembly_fragment"
133586..157263	/note="assembly_fragment"
157364..192898	/note="assembly_fragment"
192999..200734	/note="assembly_fragment"
	/note="assembly_fragment"
	clone_end:T7
	vector_side:right"
BASE COUNT	50239 a 48659 c 48384 g 51632 t 1610 others
ORIGIN	
Query Match	25.7%; Score 167.6; DB 2; Length 200734;
Best Local Similarity	63.8%; Prid.No.1.6e-23;
Matches 277; Conservative	0; Mismatch 144; Indels 13; Gaps 1;
Dy	201 CCATCAAGGCCCGAAGATCAGAAGGTGGTGGGACCTGGAAGCCGCGGGCTTTTGACT 260
Dd	145110 CCTCGAGCGCCCGAGACCAAGAAATGCGCTGGGAATTGAGAGCTCAAGGCATATGGCT 145055
Dy	261 CCAAAGCAAGTGCTACAACATCTCTGGCCGCAATGTGACGCTGGTCCGCCCATCTGCTGC 320
Dd	145050 CCAAAGCAAGTGCTACAACATCTCTGGCCGCAATGTGACATTTGGTCCCATTGGCTGC 144991
Dy	321 TCCGTGGGCTGTGTGTGTAAGTGTGACCCTGCACTGTGGCCGCTGGCCCAAGACTGTGGC 380
Dd	144990 TCCTGGGACGTGTGTGTGTAAGTGTGACCCTTGCACCTGTCCAAGTTAGCAAAACACTGTGGC 144933
Dy	381 CCTTCCTTCAAGCAAAAGTTGATGACGCGGACTATGACTGACAGGCTGGGCTCTGATCTG 440

Db	144930	CTTCTTTCAGCACCAGTTTATATAGAGAGACTATTAACCTAAGACTCCAGCTCG-----	144876
Oy	441	GGGCACTAGCCCCCAGAGACATGACCCCAAGGCTGCTGCCCTGGGGCCCAATACTACTCC	500
Oy	144875	-----TCCCTGACACGAGAGACAAACCATCTGACGTGACGTGATGCCCAACACAGCTCC	144824
Oy	501	CCGAGACCTGGGCGCTCTCTGTGTGGGGGCGTCATCCCTGACCCCATCTGATCTCGGGGGCC	560
Db	144823	TGGGAGTTACAGCTACTATAGACACTGATCTCCTGCCACCGCTGACCTCAGAGGGCGC	144764
Oy	561	CTTCAGACCCCAACATGGGACCTTAGAGCTGAACAGTCAAGCCCGGGGTCCTTACCTTA	620
Db	144763	CCCTCTGTCCACGATGTGATTTGATGTGCTGCCCACTACATCTCCTCAGACTTAATTAT	144704
Oy	621	ACCCGAGACTTCCC	634
Db	144703	AACCTTAGAGGATCC	144690
RESULT 9	AC107815	251206 bp	DNA linear
AC107815/c	Mus musculus clone RP23-114A6	WORKING DRAFT SEQUENCE, 4 ordered	
LOCUS	pieces.		
DEFINITION	AC107815	AC107815.4	GI:31442494
ACCESSION	AC107815	HTG: HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.	
VERSION	HTG: HTGS PHASE2; HTGS DRAFT; HTGS_FULLTOP.		
KEYWORDS	Mus musculus (house mouse)		
SOURCE	Mus musculus		
ORGANISM	Bkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 251206)	1	1
AUTHORS	1 Birren, B., Nusbaum, C. and Lander, E.		
TITLE	Mus musculus, clone RP23-114A6		
JOURNAL	unpublished		
REFERENCE	2 (bases 1 to 251206)		
AUTHORS	2 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Bouckigalter, B., Brown, A., Camarata, J., Campoblando, A., Chang, J., Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A., Cook, P., Deatellano, K., Dewar, K., Diaz, J. S., Dodge, S., Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Glade, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., LaRoque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McKean, P., McKernan, K., Meidrum, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, C. K., Phunhthang, P., Pierre, N., Pollara, V., Raymond, C., Retter, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.		
TITLE	Direct Submissions		
JOURNAL	Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 251206)		
AUTHORS	3 Birren, B., Nusbaum, C., Lander, E., Abouelkell, A., Allen, N., Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Bouckigalter, B., Camarata, J., Chang, J., Choepel, Y., Collymore, A., Cook, A., Cooke, P., Court, B., Deatellano, K., Diaz, J. S., Dodge, S., Doolley, K., Dorris, L., Erickson, J., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Graham, L., Grand-Pierre, N., Hafer, N., Hagopian, D., Hagos, B., Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lander, E., Levine, R., Matthews, C., Maclean, C., Lindblad-Toh, K., Liu, G., Lui, A., Mabbitt, R., Maclean, C.,		



```

/cclone="RP4-725G10"
BASE COUNT      38294 a 29779 c 30262 g 37763 t
ORIGIN
Query Match      14.7%; Score 95.4; DB 9; Length 136098;
Best Local Similarity 67.2%; Pred. No. 2.8e-09;
Matches 135; Conservative 0; Mismatches 66; Indels 0; Gaps 0;
QY 126 ACCACTGATCTGCTGGTGTTCAGACACCTCTACCTGATCTGTTGCTCGCTTCC 185
DB 68841 ACCATGTTGCTGGTCCCTGTTCAACACCTCTCATGAACCCCTGCTGCTGCTTCA 68900
QY 186 TGGGGCTGCGCTACTCTCATCAAGCCGAGATCAAGAGTGTGTGACTGGAAGCG 245
DB 68901 TAGGGTTCACTCACTCACTGAAGTCTAGGAGCAGAGATGTTGGGACCTGACCGGG 68960
QY 246 CCCGCGCTTTGGCTCCAAAGTCTCAACATCTCTGCGCGCATGTGAGCGCTGG 305
DB 68961 CCCAGGCTATGCTCTCCACCGCCAGTCTGAAACATCTGAGGCTTGACTTGGGCAATCC 69020
QY 306 TGGCGGCACCTGCTGCTCTGG 326
DB 69021 TCATGACCATCTGCTCATCG 69041

```

```

RESULT 11
AC074103/c 51891 bp DNA linear HTG 22-NOV-2000
LOCUS Homo sapiens chromosome UNK clone RP11-782B10, WORKING DRAFT
DEFINITION AC074103
SEQUENCE 3 unordered pieces.
AC074103
AC074103.4 GI:11276323
HTG: HTG_PHASE1; HTG_DRAFT.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 51891)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 51891)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Nov 22, 2000 this sequence version replaced gi:9954869.

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Project Information
Center project name: H.NH0782B10
----- Summary Statistics -----
Sequencing vector: M13; 95%
Sequencing method: Plasmid; 5%
Chemistry: Dye-terminator Big Dye; 5% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: bases at least Q40
Consensus quality: bases at least Q30
Consensus quality: bases at least Q20
Insert size: 19200; agarose-fp
Quality coverage: 4.57 in Q20 bases; agarose-fp
Quality coverage: 4.77 in Q20 bases; sum-of-contigs

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2923: contig of 2923 bp in length
2924 3023: gap of unknown length
3024 15691: contig of 12668 bp in length
15692 15791: gap of unknown length
15792 51891: contig of 36100 bp in length.
location/Qualifiers
1. 51891
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-782B10"
1. 2923
/note="assembly_name:Contig23"
3024. 15691
/note="assembly_name:Contig34"
15792. 51891
/note="assembly_name:Contig37"
BASE COUNT 13545 a 12146 c 12003 g 13995 t 202 others
ORIGIN

```

```

Query Match      14.4%; Score 93.8; DB 2; Length 51891;
Best Local Similarity 66.7%; Pred. No. 6.8e-09;
Matches 134; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 126 ACCACTGATCTGCTGGTGTTCAGACACCTCTACCTGATCTGTTGCTCGCTTCC 185
DB 19703 ACCATGTTGCTGGTCCCTGTTCAACACCTCTCATGAACCCCTGCTGCTGCTTCA 19644
QY 186 TGGGGCTGCGCTACTCTCATCAAGCCGAGATCAAGAGTGTGTGACTGGAAGCG 245
DB 19643 TACCATTCACCTCTCATCAAGTCTAGGAGCAGAGATGTTGGGACCTGACCGGG 19584
QY 246 CCCGCGCTTTGGCTCCAAAGTCTCAACATCTCTGCGCGCATGTGAGCGCTGG 305
DB 19583 CCCAGGCTATGCTCTCCACCGCCAGTCTGAAACATCTGAGGCTTGACTTGGGCAATCC 19524
QY 306 TGGCGGCACCTGCTGCTCTGG 326
DB 19523 TCATGACCATCTGCTCATCG 19503

```

```

RESULT 12
BC009696 678 bp mRNA linear PRI 22-OCT-2001
LOCUS Homo sapiens, interferon induced transmembrane protein 2 (1-8D),
DEFINITION BC009696
ACCESSION BC009696
VERSION BC009696.1 GI:16307214
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 678)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: ggaabs-remail.nih.gov
Tissue Procurement: DCTD/DRP/Gazdar
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILMIL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305

```







